Title
Mapping cellular diversity and lineage dynamics within the developing murine and human pancreas

Author
Wong, Daniel

Publication Date
2019

Peer reviewed|Thesis/dissertation

Mapping cellular diversity and lineage dynamics within the developing murine and human pancreas

by
Daniel Wong

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Todd Nystul
B55FB5CFA606452...

Todd Nystul
Chair

DocuSigned by:
Julie Beth Sneddon

Julie Beth Sneddon

DocuSigned by:
Michael German
8612790448364DD...

Michael German

Committee Members

Contributions

Experiments and analyses for this dissertation were performed in the laboratory of Dr. Julie Sneddon at the University of California, San Francisco. Chapter 2 is work published as a manuscript in *Nature Communications* with myself and Dr. Lauren Byrnes as co-first authors. Co-authors of this published manuscript include Meena Subramaniam, Nathaniel Meyer, Caroline Gilchrist, Sarah Knox, Aaron Tward, Chun J. Ye, and Julie Sneddon. I wrote this manuscript with both Drs. Lauren Byrnes and Julie Sneddon. I performed all embryonic dissections, dissociations, and FACS-sorting of cells for single-cell RNA-sequencing. I also performed immunostaining and *in situ* hybridization for single-cell RNA-sequencing validation and genetic lineage tracing experiments. Dr. Lauren Byrnes prepared single-cell RNA-seq libraries and performed all bioinformatics analyses and qRT-PCR experiments on hESC-derived beta cells. Meena Subramaniam assisted with single-cell RNA-sequencing analysis. Nathaniel Meyer performed the directed differentiation of hESCs to beta cells. Caroline Gilchrist helped perform immunofluorescence on the *Fev-Cre; Rosa26*^{mTmG} lineage traced pancreas. Human fetal pancreas was obtained from Dr. Sarah Knox. Drs. Aaron Tward and Chun J. Ye guided bioinformatics analyses. Chapter 3 contains work not yet published. Dr. Zhe Liu helped with human fetal tissue dissociation. Jacquelyn Bouza and Reed McMullen performed the directed differentiation of hESCs to beta cells. Theodore Roth guided all FEV knock-in construct designs and CRISPR/Cas9-mediated genomic editing. Gabriel Peixoto assisted in cloning of the FEV knock-in constructs and verification of clonality of all FEV knock-in and knock-out lines. I performed all remaining experiments and analyses in both chapters and wrote this dissertation.

Mapping cellular diversity and lineage dynamics within the developing murine and human pancreas

Daniel Wong

Abstract

The pancreas is a highly branched, compound gland whose development requires the diversification of many distinct cell lineages. Cells in exocrine compartment of the pancreas synthesize and secrete digestive enzymes required for food digestion, whereas cells in the endocrine compartment are responsible for maintaining glucose homeostasis through the production of various hormones. In this work, we sought to characterize the diversity of cell types present in the developing pancreas and chart their differentiation through multiple developmental stages. We start in Chapter 2 by identifying novel and known cell populations within the mesenchymal and epithelial compartments of the developing murine pancreas and mapping their developmental progress across multiple embryonic stages using single-cell RNA-sequencing. We uncover significant cellular diversity within the mesenchymal compartment and utilize single-cell transcriptomic data to reconstruct lineage relationships among the developing pancreatic mesothelium and several previously uncharacterized mesenchymal populations in mouse pancreatic development. In the epithelial compartment, we uncover a novel endocrine progenitor stage defined by high expression of a transcription factor named *Fev*. Through genetic lineage tracing, we demonstrate that this Fev+ progenitor population is derived from Ngn3+ endocrine progenitors and that all hormone-expressing endocrine lineages of the murine pancreas transit through a *Fev*-expressing cell stage. Through *in silico* reconstruction of endocrine lineages, we identify candidate regulators of alpha and beta lineage allocation expressed during this novel Fev+ progenitor stage. In Chapter 3, we apply our findings of endocrine lineage dynamics in murine pancreatic development to that of human and uncover similar *FEV*-expressing endocrine progenitor populations in human pancreatic development. We

map the transcriptional dynamics of human endocrine cell differentiation and identify novel candidate lineage regulators of human alpha and beta lineage allocation. We subsequently identify major disparities between the developmental paths that human endocrine cells follow *in vivo* versus *in vitro* during directed differentiation of hESCs towards the beta lineage. Our analysis reveals a lineage that may be mis-differentiated as hESC-derived progenitors differentiate towards a beta cell fate. Blocking the generation of this mis-differentiated lineage during *in vitro* beta cell differentiation represents a powerful solution in making *in vitro* beta cell differentiation more robust. Finally, given that FEV+ progenitors constitute a major stage in human endocrine cell differentiation, we have generated novel tools to study both the function of *FEV* and *FEV*-expressing cells in *in vitro* beta cell differentiation. Our work forms a foundation on which improvements to *in vitro* beta cell differentiation can be made to more closely reflect proper endocrine cell development *in vivo*, thereby increasing beta cell yield at the end of this process and generating beta cells that are functional.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Disease burden of diabetes mellitus

In 2017, there were an estimated 425 million people affected by diabetes mellitus in the world, representing a large global unmet need (The International Diabetes Federation). The two most prevalent types of diabetes mellitus are type 1 and type 2 diabetes (T1D and T2D). In both T1D and T2D, the function of the beta cell, the endocrine cell type housed within the pancreatic islets of Langerhans that responds to dynamic blood glucose levels by secreting insulin, is either lost or impaired. The hormone insulin is a critical signal that instructs many of the body's cells to import glucose from the blood to be used for energy (Tokarz et al., 2018). Without insulin, this regulated process of systemic energy homeostasis goes awry. In T1D, beta cells are destroyed by the host immune system, removing the only endogenous source of insulin in the body (Yoon and Jun, 2005). In T2D, insulin resistance in peripheral tissues causes beta cells to upregulate their insulin production and secretion, leading to their eventual metabolic exhaustion and the inability to meet this increased demand for insulin throughout the body (Prentki and Nolan, 2006). Fortunately, the production of insulin for human use was achieved in 1922 (Gilchrist et al., 1923). Although the administration of exogenous human insulin to patients with diabetes mellitus serves as an effective tool to manage this disease, patients with diabetes still develop life-threatening comorbidities, such as retinopathy, nephropathy, neuropathy, and macrovascular complications (Trikkalinou et al., 2017). This is because exogenous insulin administration only serves as a disease management tool and is not a curative treatment (Zheng et al., 2018). Additionally, even with exogenous insulin administration, patients with diabetes experience a diminished quality of life with the ongoing demands of diabetes management, including continually monitoring blood glucose levels, injecting insulin, abiding to a strict diet plan, carefully planning meals, and exercising (Huang et al., 2007). Thus, there is a significant need for new treatments that release patients with diabetes mellitus from dependence on exogenous insulin.

Therapeutic strategies for diabetes mellitus

Current strategies for insulin independence for patients with diabetes involve beta cell replacement, either in the form of solid pancreas or pancreatic islet transplantation (Sneddon et al., 2018). In both cases, pancreatic tissue comes from cadaveric donors, and thus, there is a limited supply of donor material that constrains these transplantation procedures from being made available to all patients with diabetes. Instead, solid pancreas or pancreatic islet transplantation is often performed as a last resort for patients with poorly controlled diabetes encompassing severe glycemic instability and worsening secondary complications such as renal failure (Giorgakis et al., 2018). The first case of insulin independence through solid pancreas transplantation was achieved in 1966 through transplantation of pancreatic tissue from a cadaveric donor into a 28-year-old recipient with T1D (Kelly et al., 1967). Successful pancreatic islet transplantation arrived in 2000 through what is now referred to as the Edmonton protocol, in which initial exogenous insulin independence was achieved in seven patients with T1D (Shapiro et al., 2000). Currently, about 50-70% of diabetic patients who undergo either islet transplantation or whole-pancreas transplantation achieve insulin independence at 5 years (Shapiro et al., 2016). Given that beta cell replacement in the form of both solid pancreas and islet transplantation has yielded favorable insulin independence outcomes for patients with diabetes, these procedures serve as proof-of-concept models in which replacement of beta cells with glucose-sensing, insulin-secreting capabilities is a clinically feasible strategy towards a therapeutic cure for patients with diabetes. However, the limited supply of whole pancreata and cadaveric islets for transplantation creates a large demand for a renewable source of transplantation material if beta cell replacement were to be achieved for all patients with diabetes.

Embryonic stem cells have the ability to differentiate into any cell type in our bodies and thus represent an attractive source of material to generate an unlimited supply of beta cells for cell replacement therapies for diabetic patients. Specifically, given that the beta cell is the cell

type that is missing or dysfunctional in diabetes, *in vitro* directed differentiation of pluripotent human embryonic stem cells into functional beta cells can address one major bottleneck in beta cell replacement strategies: the limited supply of beta cells for transplantation. One main challenge to this feat, however, is developing a robust directed differentiation program that generates glucose-sensing, insulin-secreting beta cells from human pluripotent stem cells whose function mimics that of beta cells *in vivo*. Current directed differentiation protocols for generating beta-like cells from pluripotent human embryonic stem cells (hESCs) have been developed by recapitulating many of the same developmental stages observed in pancreatic development that are required to generate beta cells *in vivo* (Pagliuca et al., 2014; Rezania et al., 2014; Russ et al., 2015). Thus, the development of these directed differentiation strategies for producing beta-like cells from hESCs has relied heavily on foundational knowledge of *in vivo* beta cell development.

Although generation of insulin-producing, beta-like cells from hESCs has been achieved, these directed differentiation strategies are not able to generate beta cells at high efficiencies or beta cells that maintain glucose-stimulated insulin secretion (GSIS) for more than 6 weeks (Pagliuca et al., 2014; Rezania et al., 2014; Russ et al., 2015; Veres et al., 2019). Overcoming these current challenges are paramount for generating pure batches of hESC-derived beta cells for transplantation into patients with diabetes for sustained exogenous insulin independence. Given that developmental biology serves as the best guide for the stereotyped development of specialized cell types *in vivo*, a more complete understanding of the proper developmental stages required to make a beta cell during pancreatic development may be needed to generate functional hESC-derived beta cells *in vitro*. Additionally, beta cell development occurs in constant interaction with other cellular compartments. Clearly annotating these cellular compartments and how they support proper beta cell development will be critical for constructing more robust directed differentiation platforms for generating hESC-derived beta cells. Therefore, additional studies on *in vivo* models of pancreatic development are needed to

continue refining our understanding of the intrinsic genetic programs and extrinsic cellular signals that are required for constructing a bona fide beta cell with proper function.

Pancreatic organogenesis: Development of the epithelial compartment

The pancreatic epithelium forms a compound gland with two functionally distinct compartments housed within the same organ. The exocrine compartment is made up of two main cell lineages: acinar and ductal. Acinar cells produce digestive enzymes required for food breakdown, and duct cells shuttle these enzymes from the pancreas into the duodenum. The endocrine compartment, on the other hand, is organized into the islets of Langerhans in which cells synthesize and secrete hormones required for energy homeostasis and nutrient metabolism. In addition to the insulin-producing beta cell, which makes up approximately 60% of the adult islet, there are four other hormone-secreting endocrine cell types that comprise the islets of Langerhans in the human adult pancreas: the glucagon-producing alpha cell (30%), somatostatin-producing delta cell (~10%), ghrelin-producing epsilon cell (~1%), and pancreatic polypeptide-producing gamma cell (less than 5%) (Da Silva Xavier, 2018). More recently, a sixth endocrine lineage, referred to as the gastrin lineage, has been described in the murine embryonic pancreas (Suissa et al., 2013). The exocrine compartment comprises the majority of the pancreatic mass (~98%), whereas islets make up ~2%. The construction of this mature pancreatic organ with distinct epithelial compartments requires the intricate coordination of cell specification, proliferation, and differentiation during development. Extensive research has focused on more clearly defining these cell lineages and the transcriptional programs that govern pancreatic organogenesis, with the hope that this knowledge can be harnessed in a therapeutic capacity for pancreatic diseases such as diabetes (Pan and Wright, 2011).

Studies of vertebrate pancreatic organogenesis have laid much of the foundation of what we understand about how pancreatic development is accomplished. Specifically in mouse,

murine pancreatic organogenesis begins at embryonic day 9 (E9) and can be segmented into two distinct periods: a primary transition that occurs E9.0-E12.5 and a secondary transition that occurs from E12.5 to birth. At the start of the primary transition, *Pancreatic And Duodenal Homeobox 1*+ (Pdx1) pancreatic progenitor cells, from which all epithelial lineages of the pancreas are derived, become specified in the dorsal and ventral foregut endoderm (Gu et al., 2003). Following specification, these Pdx1+ cells proliferate and evaginate from the foregut endoderm into the surrounding mesenchyme to form the dorsal and ventral pancreatic buds at E9.5 and E9.75, respectively, which will later fuse at approximately E11.5 due to duodenum rotation during embryogenesis (Pictet et al., 1972). These early pancreatic buds are comprised mainly of multipotent pancreatic progenitor cells (MPCs) along with early differentiated endocrine cells that express glucagon (Herrera, 2000). As these early pancreatic buds grow and evaginate during this primary transition, stratification of the epithelium leads to the formation of microlumens that will coalesce to form the tubular network reminiscent of the classic pancreatic branched morphology (Kesavan et al., 2009; Villasenor et al., 2010). Tubulogenesis and branching morphogenesis results in the spatial compartmentalization of Pdx1+ progenitors in a multipotent “tip” domain and bipotent “trunk” domain (Zhou et al., 2007). Tip cells are marked by expression of *Carboxypeptidase A1* (*Cpa1*) and *Pancreas Associated Transcription Factor 1a* (*Ptf1a*) and serve as multipotent progenitors for the ductal, acinar, and endocrine cell lineages until E13.5, after which they lose their multipotency and become restricted to the acinar fate (Zhou et al., 2007). Trunk cells, on the other hand, are limited to ductal and endocrine lineages and are marked by expression of *SRY-Box 9* (*Sox9*), *NK6 Homeobox 1* (*Nkx6.1*), and *HNF1 Homeobox B* (*HNF1β*) (Schaffer et al., 2010; Seymour et al., 2007; Solar et al., 2009). Thus, this period of development during the primary transition is responsible for expanding the pancreatic progenitor pool and structuring progenitor domains from which specialized pancreatic epithelial lineages arise.

The secondary transition of murine pancreatic development is defined by extensive differentiation and lineage allocation into the three main epithelial lineages: acinar, ductal, and endocrine. Mutual inhibitory interactions between *Ptf1a* and *Nkx6.1/Nkx6.2* transcription factors regulate lineage allocation of MPCs toward acinar versus ductal and endocrine lineages (Schaffer et al., 2010). Specifically, *Ptf1a* represses *Nkx6.1/Nkx6.2* expression to promote acinar differentiation, whereas *Nkx6.1/Nkx6.2* expression represses *Ptf1a* expression to promote ductal and endocrine fate restriction (Schaffer et al., 2010). Additionally, Notch-regulated lateral inhibition is responsible for controlling the commitment towards an endocrine or ductal cell lineage in the developing pancreatic epithelium (Apelqvist et al., 1999; Shih et al., 2012). Notch signaling promotes *Sox9* expression, which is required for both duct and endocrine differentiation (Shih et al., 2012). Cell-autonomous decreases in Notch signaling permit expression of *Neurogenin3* (*Ngn3*), a key transcription factor that demarcates endocrine progenitors that will give rise to the endocrine lineages of the pancreas, including alpha, beta, delta, epsilon, and gamma cells (Gu et al., 2002; Heller et al., 2005; Shih et al., 2012). Ngn3+ endocrine progenitors produce Notch ligands, signaling to neighboring cells to activate *Hes1*, which in return suppresses *Ngn3* expression and endocrine fate commitment. Cells with low to moderate levels of *Ngn3* fail to enter into the endocrine lineage and are directed to other pancreatic fates, including the ductal lineage in which transcription factors *Hnf1β* and *Hepatocyte nuclear factor 6* (*Hnf6*) are required (Beucher et al., 2012b; Haumaitre et al., 2005; Wang et al., 2010a; Zhang et al., 2009). Thus, high *Ngn3* expression is required for commitment to an endocrine cell fate (Beucher et al., 2012b; Wang et al., 2010a).

Pancreatic organogenesis: Transcriptional control of endocrine cell development

Given the importance of *Ngn3* expression in endocrine differentiation, extensive work has focused on the dynamics of *Ngn3* expression and the timing of differentiation of Ngn3+

progenitors throughout pancreatic development. Expression of *Ngn3* within a single cell is transient, and emergence of the Ngn3+ progenitor throughout the course of pancreatic development is biphasic, occurring in two main waves in mouse pancreatic development that encompass distinct stages of endocrine cell differentiation. The first wave of *Ngn3* expression occurs between E9.5 and E11.0 and coincides with the primary transition of pancreatic development during which the first hormone-expressing endocrine cells are observed (Villasenor et al., 2008), although these early hormone-expressing cells do not persist in the adult islet (Herrera, 2000; Herrera et al., 1998). The second wave of *Ngn3* expression begins at approximately E12.0, peaks between E13.5 and E14.5, and progressively tapers as gestation ends (Villasenor et al., 2008). This second wave of *Ngn3* expression encompasses the secondary transition of pancreatic development and produces a large fraction of hormone-expressing endocrine cells that will comprise the adult pancreas (Gu et al., 2002). Although the Ngn3+ progenitor population gives rise to the endocrine compartment of the pancreas, each individual Ngn3+ endocrine precursor has the potential to only give rise to one endocrine cell type (Desgraz and Herrera, 2009), given that Ngn3+ progenitors are post-mitotic (Miyatsuka et al., 2011). When *Ngn3* expression is induced during defined timepoints in Pdx1+ progenitors within the pancreatic epithelium using an inducible *Ngn3* expression model in a *Ngn3* null mouse, distinct endocrine lineages arise as dictated by the timing of *Ngn3* expression and are restricted to specific developmental time ranges, termed “competence windows” (Johansson et al., 2007). Ngn3+ cells that emerge at the start of pancreatic development at E9.5 differentiate almost exclusively into alpha cells, whereas Ngn3+ cells that arise starting at E10.5 have the potential to differentiate into beta and gamma lineages in addition to alpha (Johansson et al., 2007). After E14.5, the competence to form alpha cells dramatically decreases, while delta cells begin to emerge (Johansson et al., 2007). Therefore, these waves of *Ngn3* expression coupled with distinct endocrine lineage competence windows ensure proper endocrine cell differentiation throughout pancreatic development.

Coordinated with lineage allocation of Ngn3+ progenitors are the movement, rearrangement, and coalescence of differentiating endocrine progenitors to form clustered structures that will become the islets of Langerhans. *Ngn3* regulates the epithelial to mesenchymal transition (EMT) gene, *Snail2*, which represses *E-cadherin* expression in endocrine progenitors (Gouzi et al., 2011). This *Ngn3*-enforced transcriptional program is thought to facilitate endocrine progenitor exit from the pancreatic ductal epithelium during endocrine differentiation in a process known as delamination (Gouzi et al., 2011). Following delamination and subsequent endocrine differentiation, newly formed endocrine cells have been visualized to migrate and cluster with one another within the developing pancreatic structure (Puri and Hebrok, 2007), suggesting that each individual islet is not clonally derived. However, instead of outright exit of single endocrine progenitors from the pancreatic epithelium, a more recent study has proposed a model in which endocrine progenitors aggregate into peninsular structures within the epithelium and bud off in a cohesive fashion to form the endocrine clusters that give rise to adult islets (Sharon et al., 2019a). This peninsular model also involves *E-cadherin* downregulation, similar to *E-cadherin* suppression in a delamination model of endocrine development (Gouzi et al., 2011). It remains to be determined what function *Ngn3* serves in this new model of endocrine cluster formation during pancreatic development. Nevertheless, as endocrine progenitors differentiate into hormone-expressing lineages, they must also couple their differentiation to the collective clustering efforts with other proto-islet cells in order to ensure proper assembly of future mature islets.

The duration of *Ngn3* expression in a single endocrine progenitor is about 24 hours, and the timing between *Ngn3* downregulation and onset of hormone expression is estimated to be 10 hours (Beucher et al., 2012b). Despite this short duration of the Ngn3+ endocrine progenitor stage and the stages leading to acquisition of endocrine cell identity, there are several key transcriptional factors that are expressed during this period between transient Ngn3 expression and hormone expression and that are required for proper endocrine differentiation. These

transcriptional regulatory networks can be categorized into those required for proper development of all endocrine lineages and those responsible for lineage allocation into specific endocrine cell types. Transcription factors that fall into this first category include Neuronal Differentiation 1 (*NeuroD1*)*, Insulinoma-associated 1A-1 (Insm1), Regulatory factor X6 (Rfx6),* and *Islet1 (Isl1)*, all of which are transcriptional targets of *Ngn3* (Du et al., 2009; Gierl et al., 2006; Huang et al., 2000; Mellitzer et al., 2006; Naya et al., 1997; Smith et al., 2010; Soyer et al., 2010). *NeuroD1* mutant mice exhibit reduced numbers of all endocrine cell types (Huang et al., 2000; Naya et al., 1997). Like *NeuroD1* mutants, *Insm1* and *Rfx6* mutants also display are arrested at the endocrine progenitor stage and display reductions in specific endocrine lineages (Gierl et al., 2006; Mellitzer et al., 2006; Smith et al., 2010; Soyer et al., 2010). *Isl1*, on the other hand, is not required for the proper differentiation of endocrine lineages but instead is required for the proliferation and survival of differentiated endocrine cells after *Ngn3* downregulation (Du et al., 2009).

The second category of key transcriptional regulators expressed between initial Ngn3 expression and the acquisition of endocrine identity encompasses those that drive endocrine lineage-specific allocation. Following Notch-mediated lateral inhibition that specifies precursors toward the endocrine lineage through upregulation of *Ngn3*, a mutual antagonistic relationship between *Paired box 4 (Pax4)* and *Aristaless related homeobox (Arx)* transcription factors, both of which are targets of *Ngn3*, further specifies endocrine progenitors toward a beta/delta or alpha/gamma lineage (Collombat et al., 2003; Sosa-Pineda et al., 1997). *Arx* is required for alpha cell differentiation and represses acquisition of beta and delta cell identity (Collombat et al., 2003). *Pax4* functions as an antagonist to *Arx* and guides lineage allocation towards beta and delta lineage (Collombat et al., 2003; Sosa-Pineda et al., 1997). When *Arx* or *Pax4* is forcibly expressed in differentiated beta or alpha cells, respectively, *Arx*-expressing beta cells convert to an alpha identity and *Pax4*-expressing alpha cells convert to a beta lineage, demonstrating that *Arx* and *Pax4* are sufficient to drive alpha and beta lineage identities

(Collombat et al., 2007; 2009). Similar to *Arx*, *Pax6* is required for alpha cell differentiation from endocrine precursors (St-Onge et al., 1997). Additional genetic regulators of beta lineage specification include *Nkx6.1* and *Nkx2.2*. The loss of *Nkx6.1* or *Nkx2.2* blocks beta cell differentiation from endocrine precursors (Sander et al., 2000). Similar to *Pax4*, *Nkx6.1* also exhibits a mutually antagonist relationship with *Arx* to promote beta cell lineage allocation and repress acquisition of alpha cell identity (Schaffer et al., 2013). Interestingly, *Nkx2.2* mutants exhibit increased ghrelin-expressing epsilon differentiation at the expense of beta cell production, demonstrating that *Nkx2.2* promotes specific lineage allocation by suppressing genetic drivers of other endocrine lineages (Prado et al., 2004). The coordinated activities of these transcription factors during endocrine cell development ensures proper allocation of the multiple endocrine lineages from a single pool of endocrine progenitors during pancreatic organogenesis. However, a more complete understanding of how these lineage-specific transcription factors become expressed in a seemingly homogenous pool of Ngn3+ progenitors and how their expression influences when lineage specification occurs is needed.

In addition to the transcription factors expressed between *Ngn3* expression and acquisition of endocrine identity, critical transcriptional regulation after endocrine precursor differentiation ensures proper maturation of endocrine cell types and maintenance of endocrine cell identity. The majority of studies have focused on the beta cell and its maturation into a glucose-responsive, insulin-secreting cell, given the application of this knowledge to understanding and treating diabetes. *MAF BZIP Transcription Factor A (MafA)* and *MAF BZIP Transcription Factor B (MafB)* are transcription factors that activate genes required for mature beta cell function, including *insulin*, *Pdx1*, *Glut2*, and *Nkx6.1* (Artner et al., 2007; Hang and Stein, 2011; Zhao et al., 2005). During murine beta cell development *in utero*, *MafB* is expressed by newly differentiated, immature beta cells that lack GSIS capabilities (Nishimura et al., 2006). After initiation of *MafB* expression in beta cells, *MafA* becomes expressed (Matsuoka et al., 2004). Following birth, *MafB* becomes downregulated, while *MafA* expression persists in

postnatal and adult beta cells (Artner et al., 2010). Given that *MafB* is required for maintenance of key beta cell genes, *MafB* null mutants display reduced beta cell formation during embryogenesis (Nishimura et al., 2008). *MafA*, on the other hand, has no impact on beta cell formation, despite its expression during beta cell development *in utero* (Artner et al., 2007). Instead, *MafA* serves to maintain beta cell identity and promote beta cell function after the decline in *MafB* expression during early postnatal development (Wang et al., 2007). Thus, *MafA* serves as a critical regulator of beta cell maturity by activating genes associated with proper GSIS.

Human pancreatic organogenesis

Although there is extensive work on murine pancreatic development, human pancreatic development is much less systematically characterized. Human fetal development is largely divided into individual stages based on morphology and is measured in time post-conception (days post-conception (dpc) or weeks post-conception (wpc)) (O'Rahilly and Müller, 2010). Similar to that of mouse, human pancreatic development begins with demarcation of pancreatic-specified endoderm by *PDX1* expression at approximately 4wpc (Jennings et al., 2015). Multipotent progenitors in this PDX1+ region undergo a large expansion, generating a SOX9+/NKX6.1+ duct-like region reminiscent of the trunk domain in mouse pancreatic development and SOX9+/NKX6.1+/GATA4+ peripheral cells similar to mouse developmental tip cells (Jennings et al., 2015; 2013; Zhou et al., 2007). Unlike *Ngn3* expression in mouse pancreatic development, only a single, continuous wave of transient *NGN3* expression in human development starting at 8wpc has been observed, which initiates human endocrine cell development (Jennings et al., 2015; Sarkar et al., 2008). NGN3 expression peaks around 12wpc and is not detected after 35wpc (Salisbury et al., 2014). A more complete understanding

of the dynamics of critical transcription factors and their activities in human pancreatic development is direly needed.

Human islet neogenesis and structure also differ from those in mouse. The appearance of human beta cell clusters is timed with the onset of NGN3 expression at 8wpc (Jennings et al., 2013; Lyttle et al., 2008; Piper Hanley et al., 2010; Piper et al., 2004). Beta cell clusters are well vascularized by 10wpc, and aggregated islets are already composed of alpha, beta, delta, and gamma cells by 12-13wpc (Jennings et al., 2013; Piper et al., 2004). In mouse pancreatic development, preliminary islet formation occurs late in gestation and continues into postnatal development. As proper islet structure takes shape, human islets display a comingling of alpha and beta cells throughout their structure, whereas mouse islets adopt a core-mantle structure in which alpha cells are located at the periphery and beta cells largely comprise the islet core (Kim et al., 2009). Understanding these structural differences that arise during mouse and human islet neogenesis and their functional outcomes will be critical for optimizing current *in vitro* efforts that attempt to construct 3D islets from stem cell-derived sources.

The role of the microenvironment in shaping pancreatic epithelial development

The epithelial compartment in the developing pancreas is in close association with the surrounding microenvironment, and these interactions have critical roles at defined times in pancreatic organogenesis for proper epithelial development. The earliest of such known interactions involves contact with various embryonic structures that orchestrate the process of development of both the dorsal and ventral pancreas, which later fuse to give rise to a singular solid organ (Pan and Wright, 2011). As early at E8, the notochord is in contact with the prospective dorsal pancreatic endoderm and is responsible for repressing *Sonic Hedgehog* (*Shh*) in the pancreatic endoderm through the secretion of activin-betaB and *Fibroblast growth factor 2* (Fgf2) (Hebrok et al., 1998). This suppression of *Shh* ensures proper morphogenesis

and permits subsequent *Pdx1* induction in the pancreatic endoderm (Hebrok et al., 1998). The induction of *Pdx1* in the dorsal endoderm is accomplished by contact with the dorsal aorta, which displaces the notochord by E8.5 (Lammert et al., 2001; Slack, 1995; Yoshitomi and Zaret, 2004). Similar to dorsal pancreatic development, in ventral pancreatic development, contact with vitelline veins of the developing vasculature induces a pancreatic transcriptional program that includes *Pdx1* and *Ptf1a* expression in the prospective ventral pancreatic endoderm (Lammert et al., 2001; Yoshitomi and Zaret, 2004). Without these intricate, step-wise physical interactions with various embryonic structures, specification of the pancreatic epithelium from the foregut endoderm does not occur (Hebrok et al., 1998; Lammert et al., 2001; Yoshitomi and Zaret, 2004).

Following pancreatic bud specification, many cell types in the microenvironment work together to support proper growth, differentiation, and maturation of the pancreatic epithelium into a highly branched, 3D network. These include the vasculature, nerves, and mesenchymal cells. After serving to first promote early pancreatic specification, the developing vasculature plays different roles in the development of the epithelial compartment. In hypervascularization models of pancreatic development, exocrine differentiation is significantly reduced (Magenheim et al., 2011; Pierreux et al., 2010; Sand et al., 2011). Conversely, inhibition of blood vessel vascularization through small molecule treatment in the developing pancreas results in increased exocrine cell differentiation and a reduction in endocrine cell differentiation (Pierreux et al., 2010). Together, these findings suggest that endothelial cells serve to restrict exocrine lineage differentiation and promote endocrine cell differentiation. These changing roles of the vasculature could be the result of dynamic interactions that blood vessels have with the epithelial compartment as they invade and intercalate between epithelial branches (Villasenor and Cleaver, 2012). Thus, the varying interface between the vasculature and the pancreatic epithelium at different times in development may influence its dynamic functions in pancreatic organogenesis.

Additionally, pancreatic development is dependent on proper sympathetic innervation, which is required not only for proper formation of the islets of Langerhans but also for their maturation into functional, glucose-sensing units (Borden et al., 2013). Both genetic and pharmacologic inhibition of sympathetic innervation in murine pancreatic development inhibits islet cell migration that facilitates formation of coalesced islets, demonstrating that the endocrine compartment relies on guidance cues from sympathetic nerves to establish correct islet architecture (Borden et al., 2013). One such cue is norepinephrine, which activates beta-adrenergic signaling in migrating islet cells to enforce proper islet architecture (Borden et al., 2013). These resulting defects in islet architecture from inhibition of sympathetic innervation lead to reduced insulin secretion and impaired glucose tolerance (Borden et al., 2013), highlighting the functional consequences of the absence of sympathetic nerves in the microenvironment that surrounds the developing epithelium during pancreatic organogenesis.

Finally, the pancreatic mesenchyme has also been demonstrated to support the development of the pancreatic epithelium. Early studies of pancreatic development showed that physical removal of the mesenchyme in *ex vivo* culture blocked proper growth and morphogenesis of the pancreatic epithelial buds cultured *ex vivo* (Golosow and Grobstein, 1962). These findings were later corroborated using more specific genetic ablation methods, where genetic ablation of the pancreatic mesenchyme *in vivo* halted expansion of both the endocrine and exocrine compartments of the pancreatic epithelium (Landsman et al., 2011). Fgf10, Wnt5a, and BMPs are all secreted factors expressed by the pancreatic mesenchyme whose deletion from the mesenchyme leads to reduced epithelial growth, branching, and differentiation (Ahnfelt-Rønne et al., 2010; Bhushan et al., 2001; Larsen et al., 2015). Beyond its role as a source of signaling cues, the mesenchyme also produces many ECM components that may be structurally important for the developing epithelium. The proportion of epithelium to mesenchyme steadily increases as the epithelium grows into a highly-branched structure with mesenchyme nestled between epithelial branches by the end of gestation. These direct,

physical interactions that the mesenchyme has with the epithelium may be important for supporting physical growth of the branched organ.

Although the importance of the pancreatic mesenchyme on proper epithelial development has been established, there is not a well-established definition of what the pancreatic mesenchyme is or which cell types constitute this compartment. Classically, mesenchyme is defined as embryonic connective tissue made of loosely associated, fibroblast-shaped cells of mesodermal origin (Hay, 2005). Defining pancreatic mesenchymal cells has largely depended on morphological features, such as elongated shape. Few molecular markers have been identified to specifically mark this population in pancreatic development, although genes including those in the collagen family, *Pdgfra*, and *vimentin* have been used. Exclusion of known epithelial markers, such as *Epcam* and *E-cadherin* has also been utilized as a strategy to define cells of the pancreatic mesenchymal compartment. Given the poor annotation of the entire mesenchymal compartment within pancreatic organogenesis, a better cellular and molecular atlas of the pancreatic mesenchyme is direly needed. In particular, one key question is whether the pancreatic mesenchyme is composed on a singular cell type or if multiple cell populations with distinct functions exist within this compartment. Clarifying lineage relationships of cells within the pancreatic mesenchyme will also enable better characterization of this compartment as a whole. By defining molecular markers of the different cell populations within the mesenchyme, more precise targeting strategies to study the different functions of different mesenchymal populations can be developed.

Directed differentiation of human embryonic stem cell-derived beta cells

The key development cues found in pancreatic development laid the roadmap through which hESCs are differentiated through a step-wise process, first to definitive endoderm, then to a pancreatic progenitor stage, followed by transit through an endocrine progenitor stage, and

finally terminating in insulin-expressing cells of the beta lineage (Pagliuca et al., 2014; Rezania et al., 2014; Russ et al., 2015). The first hurdle in the development of directed differentiation protocols to generate hESC-derived beta cells was to induce the definitive endoderm lineage from hESCs (D'Amour et al., 2005). It is known *in vivo* that Nodal/activin signaling induces endoderm during gastrulation in vertebrates (Stainier, 2002). The addition of Activin A to hESC culture successfully differentiated hESCs toward the FOXA2+/SOX17+ definitive endoderm lineage (D'Amour et al., 2005). Following the production of definitive endoderm, generation of PDX1+ pancreatic progenitors was accomplished through the addition of retinoic acid, FGFs, and hedgehog inhibitors (D'Amour et al., 2006), all of which facilitate entry into the pancreatic lineage (Lau et al., 2006) and derive from the adjacent mesoderm or notochord during pancreatic development (Hebrok et al., 1998). In pancreatic development, Pdx1+ progenitors are multipotent and differentiate into acinar, ductal, and endocrine lineage (Zhou et al., 2007). The endocrine lineage begins with the onset of Ngn3 expression, which is induced with Notch inhibition (Apelqvist et al., 1999). Therefore, successful endocrine cell specification followed by the production of hESC-derived, PDX1+ pancreatic progenitors was accomplished through use of Notch inhibitors to permit expression of the pro-endocrine transcription factor *NGN3*, leading to the production of endocrine progenitors (D'Amour et al., 2006). From these NGN3+ endocrine progenitors, several groups have gone on to generate glucose-sensitive, insulin-producing beta cells (Nair et al., 2019; Pagliuca et al., 2014; Rezania et al., 2014; Russ et al., 2015; Veres et al., 2019). The studies on pancreatic development that have identified genetic programs and signaling pathways required for transit of pluripotent cells and progenitors into the pancreatic and subsequent endocrine lineages *in vivo* thus have been instrumental in the derivation of step-wise differentiation protocols that generate beta cells from hESCs.

Although the successful production of glucose-sensitive, insulin-producing beta cells was a major leap forward, there are several key limitations that remain before hESC-derived beta cells can be used as a therapeutic intervention for diabetes. First, the number of beta cells

required for transplantation into one diabetic patient is on the order of one billion (Jacobson and Tzanakakis, 2017; Lock and Tzanakakis, 2007). Although current hESC-derived beta cell differentiation protocols are capable of differentiating millions of hESCs at a time, the efficiency of generating beta cells is batch-dependent and low, at approximately 30-40% (Pagliuca et al., 2014; Rezania et al., 2014; Russ et al., 2015) and even reported purification methods are low-throughput and labor-intensive (Nair et al., 2019; Veres et al., 2019). Given the number of hESC-derived beta cells required to treat diabetes in one patient, improvements in the efficiency of functional beta cell production are needed for these regenerative-based therapies to be scalable and of consistent quality. Second, one key hallmark of a functional beta cell is that it can properly sense dynamic glucose levels and secrete insulin in a glucose concentration-dependent manner (Blum et al., 2012). Although the hESC-derived beta cells from recent directed differentiation protocols initially exhibit glucose-sensitive insulin secretion (GSIS) (Pagliuca et al., 2014; Rezania et al., 2014; Russ et al., 2015), they are not able to maintain GSIS long-term. This lack of functionality of hESC-derived beta cells long-term makes them unsuitable as transplantation material for diabetes treatment. Third, although beta cells in human pancreatic islets have been thought to be a homogenous cell population, the level of beta cell heterogeneity in human islets has recently begun to be appreciated. Based on two cell surface markers alone, ST8SIA1 and CD9, four antigenically distinct beta cell subpopulations have been identified in human islets, and the frequencies of these beta cell subtypes are altered in individuals with T2D (Dorrell et al., 2016). Specialized beta cells with pacemaker properties residing in islets, deemed “hub cells,” have also been identified and regulate the coordinated islet responses to glucose (Johnston et al., 2016), supporting this notion that not all beta cells are created equal in the islet. Thus, future differentiation strategies should focus not only on the generation of functional beta cells but also on the formation of beta cell subtypes that may significant for the overall function of an islet. Lastly, stem-cell derived beta cells generated *in vitro* must be protected against immune attack in transplant recipients. Several strategies have

been formulated to circumvent this issue of immune system attack, including encapsulating hESC-derived beta cells in an immunoprotective device (Basta et al., 2011; Sneddon et al., 2018), genetic engineering of hESC-derived beta cells to resist immune detection (Sneddon et al., 2018), and generating beta cells from human induced pluripotent stem cells (hiPSCs) derived from the transplant recipient (Millman et al., 2016). Addressing these limitations will prove critical for the adoption of stem-cell derived beta cells for use in the clinic to address diabetes.

These challenges in beta cell functionality and differentiation efficiency in hESC-derived beta cell differentiation protocols can be addressed by better understanding the gene expression programs that drive beta cell differentiation and maintain beta cell identity *in vivo*. Given that beta cells are tasked with the critical responsibility of producing, storing, and releasing insulin at dynamic glucose levels, beta cells require a host of transcription factors that activate expression of proteins that mediate beta cell function. Although current beta cell differentiation methods achieve expression of key transcription factors such as *PDX1, NKX6.1, NGN3*, and *NEUROD1* (Pagliuca et al., 2014; Rezania et al., 2014; Russ et al., 2015), these methods may be failing to activate other transcriptional programs not yet uncovered to be required for proper beta cell differentiation and function. Additionally, although current differentiation protocols for generating hESC-derived beta cells recapitulate key developmental stages and genetic programs used to make beta cells *in vivo*, we may not be fully mimicking the exact developmental path taken *in vivo* in *in vitro* approaches. For instance, incorporation of endothelial cells into insulin-producing cell implantation promotes functional maturation of these insulin-producing cells (Meivar-Levy et al., 2019). It is possible that there are several different differentiation trajectories that exist to push cells into the beta lineage, but only a select few generate the functional, bona fide beta cells observed *in vivo*. Current differentiation protocols span several weeks, but beta cell development in human gestation occurs over months. Thus, the compressed window of beta cell development *in vitro* in differentiation protocols may prevent

the generation of fully functional hESC-derived beta cells. Onset, duration, and dosage of expression of key genes must be finely coordinated during beta cell development, and these may not be fully refined in directed differentiation protocols to yield fully functional beta cells.

Additionally, current differentiation strategies that result in poor differentiation efficiency and reduced beta cell functionality may require the support of other cell types known to form the microenvironment during pancreatic development. Although extensive work has elucidated several key cell intrinsic programs activated and repressed as progenitors differentiate into beta cells, this process of endocrine differentiation does not occur in isolation *in vivo*. Instead, beta cell differentiation and proper function are carried out in the presence of a milieu of other cell types, including other hormone-expressing islet cells, endothelial cells, perivascular cells, nerves, and mesenchymal cells (Hayden et al., 2008; Lammert et al., 2003; Reinert et al., 2014). These additional cell types may serve important niche functions for supporting beta cell differentiation and function, and the lack of these cellular interactions in current *in vitro* beta cell differentiation platforms directly contrasts that seen in the *in vivo* microenvironment. Proper recapitulation of the microenvironment during *in vitro* derivation of beta cells may enforce proper beta cell differentiation and function (Sneddon et al., 2018). When hESC-derived pancreatic progenitor cells are transplanted in mice, these progenitors differentiate and mature into glucose-responsive insulin-producing beta cells (Bruin et al., 2015; Sneddon et al., 2018), suggesting that additional components found in the cellular microenvironment can facilitate proper beta cell differentiation and enforce proper GSIS function. Additionally, co-culture of pancreatic mesenchyme enhances proliferation and self-renewal of hESC-derived pancreatic progenitors that go on to give rise insulin-secreting beta cells (Sneddon et al., 2012), highlighting the beneficial aspects of including cells from the endogenous microenvironment in the directed differentiation of hESC-derived beta cells. Incorporation of endothelial cells also improved function of transplanted islets (Penko et al., 2011), suggesting that endothelial cells, too, can improve the functionality of transplanted hESC-derived beta cells. A more defined map

of the intricate cellular interactions within the microenvironment during beta cell differentiation and function *in vivo* will enable the development of improved *in vitro* beta cell differentiation platforms with better outcomes in terms of differentiation efficiency and beta cell functionality.

Contribution to the field

This work builds a transcriptomic map of cellular dynamics and lineage relationships in both murine and human pancreatic development. In the developing murine pancreas, we use single-cell RNA-sequencing to characterize the previously underappreciated diversity of cell types present in the mesenchyme and construct lineage relationships among the pancreatic mesothelium and newly annotated mesenchymal populations. We apply a similar approach to the epithelial compartment and provide increased resolution to the cellular stages involved in endocrine cell differentiation of the pancreas. We identify a novel endocrine progenitor stage defined by differential expression of a transcription factor named *Fev* and show that Fev+ endocrine progenitors are derived from Ngn3+ progenitors and give rise to hormone-expressing lineages of the murine pancreas. By combining genetic lineage tracing of *Fev*-expressing cells with single-cell RNA-sequencing, we construct a high-resolution transcriptomic map of endocrine cell differentiation into the alpha and beta lineages and identify novel candidate regulators of alpha and beta fate restriction. Thus, our transcriptional profiling will serve as a rich resource for further understanding the cellular diversity present in both the developing mesenchymal and epithelial compartments of the murine pancreas.

We extend our work in murine pancreatic development to examine the cellular heterogeneity present in human fetal pancreas through single-cell transcriptomic profiling. Examination of the human endocrine compartment of the fetal pancreas reveals a clearer transcriptomic profile of NGN3+ endocrine progenitors and identifies pre-alpha and pre-beta progenitors that express *FEV*. *In silico* reconstruction of lineage relationships in the alpha and

beta lineages uncover novel candidate regulators of lineage allocation. Using our single-cell transcriptomic profiling of human pancreatic development as a guidepost, we build a map of the cellular dynamics present during the directed differentiation of hESC-derived beta cells. This map of *in vitro* beta cell differentiation uncovers a novel lineage that results from mis-differentiation of FEV-expressing progenitors and opens avenues through which current *in vitro* beta cell differentiation methods can be improved for greater differentiation efficiency. Given our findings of Fev/FEV+ progenitors in murine development, human fetal development, and the *in vitro* derivation of beta cells, we engineer valuable hESC lines to study the function of the *FEV* gene and FEV+ endocrine progenitors during the directed differentiation of beta cells and possibly, in the future, other hormone-expressing endocrine lineages. This work constructs a new differentiation model for human endocrine cell development and paves the way for improved *in vitro* beta cell derivation methods that better reflect *in vivo* human beta cell development.

Chapter 2

Lineage dynamics of murine pancreatic development at single-cell resolution

Introduction

Pancreatic organogenesis is a complex and dynamic process that ultimately results in the generation of multiple cell lineages that perform the functions of the mature organ: the regulation of glucose homeostasis by the endocrine compartment and the production of digestive enzymes by the exocrine compartment. In the mouse, all known epithelial lineages of the pancreas derive from a small field of epithelial precursor cells within the foregut endoderm specified by the expression of *Pancreatic duodenal transcription factor 1* (*Pdx1*) (Fig. 2.1a) (Shih et al., 2013). These Pdx1+ cells evaginate into a cap of surrounding mesenchymal cells around embryonic day 9 (E9), proliferate, and begin the process of branching morphogenesis. Further epithelial lineage diversification continues with the specification of Pdx1+ cells into tip and trunk domains by E12 and progresses to the restriction of tip cells to a digestive enzyme-producing acinar fate and of trunk cells to either a ductal or endocrine cell fate (Shih et al., 2013). Within the trunk domain, induction of *Neurogenin 3* (*Ngn3*) expression defines the cells that will differentiate into one of five endocrine lineages: alpha, beta, delta, gamma, or epsilon cells, marked by expression of the hormones Glucagon (Gcg), Insulin (Ins), Somatostatin (Sst), Pancreatic polypeptide (PP), or Ghrelin (Ghrl), respectively (Pan and Wright, 2011). Gastrin+ cells have also been recently described (Suissa et al., 2013). Despite previous work focused on the formation of the endocrine compartment, the precise timing and coordination of lineage decisions are not completely understood.

Although the pancreatic mesenchyme is required for the proper differentiation, proliferation, and morphogenesis of the epithelial network (Shih et al., 2013), little is known about the cell identities and lineages that compose the pancreatic mesenchyme during development. Even less is known about the mechanisms by which these distinct mesenchymal cell types interact with one another and with the cells of the epithelial compartment during development and in the adult organ. Therefore, a deeper understanding of the full diversity of

the mesenchymal cell types, as well as their global gene expression profiles, will serve as a basis for understanding these key cellular interactions.

Recent studies of late embryonic, postnatal, and adult alpha and beta cells have demonstrated the power of single-cell transcriptomic profiling for unraveling endocrine lineage heterogeneity and revealing distinct transcriptional states of beta cell maturation (Dorrell et al., 2016; Qiu et al., 2017a; Zeng et al., 2017). Here, we perform droplet-based, single-cell RNA-sequencing of entire murine embryonic pancreata at earlier developmental timepoints to describe the cellular diversity and dynamics of gene expression in both the epithelial and mesenchymal compartments. We further identify and validate these populations within mouse and human pancreatic tissue, as well as human embryonic stem cell (hESC)-derived endocrine progenitor cells. Finally, we predict lineage relationships, identify previously unappreciated intermediate progenitor cells, and validate our methodology using *in vivo* genetic lineage tracing.

Results

Cellular heterogeneity in the murine pancreas

We first characterized the major sources of cellular heterogeneity in the developing pancreas. Two batches of mouse pancreata at E14.5, a particularly active time of expansion, morphogenesis, and diversification (Pan and Wright, 2011) (Fig. 2.1a), were dissected from individual litters, dissociated into single-cell suspensions, sorted for live cells, and sequenced using the 10X Chromium Single-Cell version 1 (v1) kits (Fig. 2.1b and Fig. 2.2a-e). We performed filtering, normalization, variable gene identification, linear regression for batch, and Principal Component Analysis (PCA) with the R package, Seurat (Fig. 2.2d,e and 3a,b). Graph-based clustering (Satija et al., 2015) of batch-adjusted, merged data identified 19 distinct cell populations, classified as epithelial, mesenchymal, immune, or vascular populations based on

the expression of known markers (Fig. 2.1c,d and Supplementary Data 1). We identified expected populations, including endocrine, exocrine (acinar and ductal), and endothelial cells (Fig. 2.1e). The proportions of endocrine, mesenchymal, immune, and vascular populations were similar between E14.5 batches (Fig. 2.3b-d). Downsampling analysis confirmed that sufficient sequencing depth had been reached for calling clusters (Fig. 2.3e-g). These results reveal the power of single-cell RNA-sequencing to identify a broad range of cell types during development.

Characterization of mesenchymal heterogeneity

While previous studies have identified numerous markers of pancreatic epithelial populations(Pan and Wright, 2011), comparatively little is known about heterogeneity among pancreatic mesenchymal cells. We characterized the mesenchymal compartment by sub-clustering only mesenchymal cells (5,069 cells) and re-performing the clustering analysis (Fig. 2.4a and Fig. 2.5a). Despite being less divergent from one another than were cells in the epithelial compartment (Fig. 2.4b and Fig. 2.5b), mesenchymal cells could still be sub-divided into 10 transcriptionally distinct mesenchymal clusters (Fig. 2.4a,c and Supplementary Data 2). We verified the differential gene expression analysis with three tests: bimodal likelihood ratio test (McDavid et al., 2012), Wilcoxon rank sum, and MAST (Finak et al., 2015) (Fig. 2.5c and Supplementary Data 2). We annotated two clusters based on the expression of known marker genes: cluster 1 is pancreatic mesothelial cells (*Wt1*, *Krt19*, and *Upk3b*) (Kanamori-Katayama et al., 2011; Winters and Bader, 2013) and cluster 3 represents vascular smooth muscle (VSM) cells (*Acta2*, *Tagln*, and *Myl9*) (Fig. 2.4c and Supplementary Data 2) (Majesky et al., 2011). Indeed, in E14.5 pancreas, Wt1 expression was restricted to the tissue edge, as expected for mesothelial cells, while Acta2 expression was localized to cells surrounding vessels, as expected for VSM cells (Fig. 2.5d,e). Cells in the mesothelial cluster also expressed the

secreted factors *Fgf9*, *Pdgfc*, *Rspo1, and Igfbp5* (Fig. 2.5f) and genes regulating prostaglandin hormone signaling and tight junctions (Fig. 2.4d and Supplementary Data 3).

The remaining mesenchymal clusters included proliferating cells (clusters 6, 7, and 8), a large cluster (10) expressing pan-mesenchymal markers, and four clusters (2, 4, 5, and 9) each expressing a signature distinct from that of cluster 10 (Fig. 2.4a,c and Supplementary Data 2). Cluster 2 was defined by differential expression of *Stathmin 2* (*Stmn2*), a gene involved in neurite outgrowth and osteogenesis (Chiellini et al., 2008; Grenningloh et al., 2003). We also found two populations, clusters 4 and 5, that differentially expressed multiple secreted factors. Cluster 4 expressed *Ace2*, the chemokines *Cxcl12* and *Cxcl13*, and *Vegfd*, while cluster 5 expressed high levels of the Wnt antagonists *Secreted frizzled-related protein 1 and 2* (*Sfrp1* and *Sfrp2*) (Fig. 2.4c-e and Supplementary Data 2). Cluster 5 also expressed the transcription factor *BarH-like homeobox1 1* (*Barx1*) and members of the Id DNA-binding protein family (Fig. 2.4c-e and Supplementary Data 2). Cluster 9 expressed *Nk2 homeobox 5* (*Nkx2-5*) and *Tlx1*, transcription factors reported to contribute to splenic development during a window in which the embryonic pancreas and spleen share a mesenchymal compartment (Fig. 2.4c) (Hecksher-Sørensen et al., 2004). Pathway analysis identifies multiple signaling pathways that may be functionally relevant in these populations (Fig. 2.4d and Supplementary Data 3). We validated a subset of these distinct clusters using dual *in situ* hybridization/immunofluorescence (ISH/IF) on E14.5 pancreas for differentially-expressed markers of clusters 1 (*Cav1* and *Barx1*), 2 (*Stmn2*), and 5 (*Barx1*) (Fig. 2.4e-h). These gene expression profiles demonstrate a previously underappreciated level of heterogeneity in the mesenchymal compartment of the developing pancreas.

Mesothelial cells undergo changes across developmental time

During organogenesis, the dynamics of each lineage are defined by the expansion, differentiation, and maturation of its constituent cells. To address how these processes change across chronological time within the developing pancreas, we performed single-cell sequencing at two additional timepoints, E12.5 and E17.5 (Fig. 2.6a). We identified mesenchymal cells from E12.5, E14.5, and E17.5 timepoints, merged them into one dataset, and re-performed the clustering analysis. We identified the clusters detected in our E14.5 analysis (clusters 1-10) along with seven new clusters (11-17) (Fig. 2.6a and Supplementary Data 4). The addition of E12.5 and E17.5 cells revealed further sub-division of the mesothelium into timepoint-specific clusters (1, 11, and 17), each with unique transcriptomic signatures (Fig. 2.6a,b). Within the mesothelium, we verified *Paired-like homeodomain transcription factor 2* (*Pitx2*) expression at E12.5 and its absence at E17.5 and *Mesothelin* (*Msln*) expression at E17.5 and its absence at E12.5 (Fig. 2.6c), consistent with the single-cell data. These data provide evidence of transcriptional maturation over developmental time within the mesothelial compartment.

While the mesothelium is a well-established mesenchymal progenitor cell population for VSM and fibroblasts in multiple other organs, both the role of the mesothelium and the origin of the mesenchymal cell types within the pancreas remain uncharacterized (Wilm et al., 2005). We utilized our single-cell mesenchymal dataset to determine whether the pancreatic mesothelium may function as a mesenchymal progenitor cell population during development. We found six populations (clusters 2, 3, 4, 5, 12, and 13) that expressed VSM cell genes, such as *Acta2* and *Tagln*, or genes known to regulate VSM development, such as *Mgp* (Speer et al., 2009), *Fhl1* (Kwapiszewska et al., 2008; Wang et al.), *Barx1* (Jayewickreme and Shivdasani, 2015), and *Pitx2* (Shang et al., 2008) (Fig. 2.6d). Based on these VSM-related gene expression profiles, we hypothesized that these populations could represent VSM progenitors derived from the pancreatic mesothelium. To test the lineage relationships among these populations, we ordered cells in pseudotime based on their transcriptional similarity (Qiu et al., 2017b). This analysis

placed mesothelial cells on one side of the pseudotime trajectory (Fig. 2.6e). Mesothelial branches corresponded to either a maturation process, based on placement of E17.5 cells at the branch terminus, or proliferating mesothelium, based on expression of proliferation genes (Fig. 2.6e). VSM-related populations were placed on the other side of the trajectory (Fig. 2.6e). We calculated the proportion of each population over pseudotime and found a transition from the E12.5 mesothelial population (cluster 11) to cluster 12, both of which share expression of the gene *Pitx2* (Fig. 2.6e-g). Cluster 12 then transitioned into the *Stmn2*-expressing cluster 2, which split into a branch composed of VSM populations, clusters 3 and 13 (Branch 1), and a branch composed of clusters 4 and 5 (Branch 2) (Fig. 2.6e-g). Thus, this analysis proposes clusters 2 and 12 as potential mesothelial-derived mesenchymal progenitor populations that can contribute to the VSM lineages (Fig. 2.6g). Our analysis has identified and validated multiple mesenchymal subtypes and possible lineage relationships among them.

A previously undescribed endocrine progenitor population

After assessing the heterogeneity within the mesenchymal compartment, we next focused on the epithelial cells. We first sub-clustered the 2,049 cells from our E14.5 dataset that comprised just the epithelial populations (Fig. 2.7a and Fig. 2.8a). We identified 10 clusters, including acinar, ductal, beta, alpha, and Ngn3+ progenitor populations, as revealed by differential expression of known markers (Fig. 2.7a,b, Fig. 2.8b, and Supplementary Data 5). Our analysis highlighted previously uncharacterized markers of acinar, Ngn3+, beta, and alpha cell populations, such as *Reep5*, *Btbd17*, *Gng12,* and *Peg10,* respectively (Fig. 2.7b and Supplementary Data 5). We also found *Sst*- and *Pancreatic polypeptide* (*Ppy*)-expressing cells, but they did not cluster into their own populations (Fig. 2.8c).

After the ductal, acinar, Ngn3+, and hormone+ populations had been accounted for, there still remained one population that eluded classification based on known marker genes.

This population was distinguished from all other epithelial populations by high-level expression of the E26 transformation-specific (ETS) transcription factor *Fev*, previously shown to be expressed within the developing pancreas but not described as a marker of a distinct epithelial population (Ohta et al., 2011) (Fig. 2.7a,b and Supplementary Data 5). This Fev+ population expressed genes marking endocrine lineage cells, such as *Paired box 4* (*Pax4*), *chromogranins A/B (Chga/b)* and *Neurod1* (Shih et al., 2013) (Fig. 2.8d), but not mature endocrine markers, such as *Insulin1* (*Ins1*) or *Gcg*, or the transitory early endocrine lineage marker, *Ngn3* (Fig. 2.7b,c and Supplementary Data 5). Pairwise comparison between the Fev+ and Ngn3+ clusters identified 99 genes more highly expressed in Fev+ and 87 more highly expressed in Ngn3+ cells, suggesting that they are distinct populations (Fig. 2.7d). This Fev+, Ngn3-, hormone- cluster will henceforth be referred to as the Fev^{Hi} population. Pathway analysis of the Ngn3+ and Fev^{Hi} populations revealed enrichment of cell cycle and Notch signaling pathways in Ngn3+ cells (Fig. 2.7e and Supplementary Data 6), likely reflecting the exit of Ngn3+ progenitors from the cell cycle (Miyatsuka et al., 2011) and the role of *Ngn3* in Notch signaling (Shih et al., 2012). The Fev^{Hi} cluster expressed genes in pathways related to serotonin and insulin signaling, Activating Transcriptional Factor 2 (ATF-2) signaling, and sphingosine-1-phosphate signaling, which have been reported to regulate endocrine differentiation (Han et al., 2011; Serafimidis et al., 2017). This relationship to serotonin is consistent with prior work establishing *Fev* as a critical transcription factor in serotonergic neurons (Ohta et al., 2011; Spencer and Deneris, 2017).

Further sub-clustering of all 661 cells within the endocrine lineage revealed additional sub-groups of *Fev*-expressing cells. The first was marked by high expression of *Pax4* and *Runx1 Translocation Partner 1* (*Runx1t1*) and lower levels of *Ngn3*. The second was marked by *Chgb* and *Vimentin* (*Vim*) (Fig. 2.7f, Fig. 2.8e,f, and Supplementary Data 7). Therefore, our analysis proposed the existence of multiple intermediate states, marked by *Fev*, within the endocrine lineage. The *Fev* gene was also expressed at lower levels in a subset of the

hormone-producing alpha, beta, and epsilon cell populations, which will collectively be referred to as hormone+/Fev^{Lo} populations (Fig. 2.7b).

Given that the Fev+ populations expressed endocrine lineage genes, we utilized pseudotime ordering (Qiu et al., 2017b) to test the hypothesis that both Fev+ populations were lineage-related to the Ngn3+ progenitors that give rise to the endocrine compartment of the pancreas (Gu et al., 2002). This *de novo* reconstruction of the developmental trajectory placed both the Fev+/Pax4+ and Fev^{Hi}/Chgb+ cells between Ngn3+ endocrine progenitors and alpha and beta cells (Fig. 2.7g), suggesting that Fev^{Hi} cells comprise a progenitor stage following *Ngn3* expression and before hormone acquisition. The Fev+/Pax4+ population was placed closer in pseudotime to the Ngn3+ population and was followed by the Fev^{Hi}/Chgb+ population (Fig. 2.7g), indicating that the former represents an earlier cell state. Unlike alpha and beta cells, epsilon cells were found throughout the trajectory populated by the Fev+/Pax4+ and Fev^{Hi}/Chgb+ populations (Fig. 2.7g), possibly reflecting their function as multipotent progenitor cells for alpha and gamma lineages during development (Arnes et al., 2012).

To validate these lineage relationships, we performed an *in vivo* lineage trace of Ngn3+ cells. In E14.5 *Ngn3-Cre; ROSA26^{mTmG}* mouse pancreata, where lineage-traced cells are membrane-GFP+ (Muzumdar et al., 2007), approximately 20% of all *Ngn3*-lineage traced cells were identified as the Fev^{Hi} population by the presence of *Fev* and the absence of both *Ngn3* and the pan-differentiated endocrine cell marker *Islet1* (*Isl1*) (Fig. 2.9a,e, yellow arrows and bar). We also detected the hormone+/Fev^{Lo} population identified by our single-cell data (Fig. 2.9a, purple arrows) and cells that co-expressed *Fev* and *Ngn3* (blue arrows), consistent with a model in which Fev^{Hi} cells represent an intermediate progenitor state following Ngn3+ cells but prior to differentiated endocrine cells (Fig. 2.9g).

We next tested if the Fev^{Hi} population was also present in developing human pancreatic tissue. In human fetal pancreas at 23 weeks post conception, we observed cells that expressed only *NGN3* (Fig. 2.9b, grey arrows), only CHGA (magenta arrows), a marker of all hormone-

expressing endocrine cells, and both *FEV* and CHGA (purple arrows). We also detected cells that expressed *FEV* but not *NGN3* or CHGA (Fig. 2.9c, yellow arrows), analogous to the murine Fev^{Hi} population. The existence of these cellular states in human development suggests that the lineage relationships we identified generalize beyond murine pancreatic organogenesis to that of human, as well.

We then probed hESCs undergoing directed differentiation towards the pancreatic beta cell lineage *in vitro* (Pagliuca et al., 2014). *FEV* was detected in endocrine progenitor-stage cells and beta-like cells (BLCs) at levels comparable to adult human islets, but not in undifferentiated hESCs (Fig. 2.8i). Further, we observed *FEV*+ (*NGN3*-/*ISL1*-) (yellow arrows), *FEV*+/*ISL1*+ (*NGN3*-) (purple arrows), and *NGN3*+/*FEV*+ (*ISL1*-) (blue arrows) populations in differentiating hESC-derived cells mid-way through the endocrine progenitor stage (Fig. 2.9d,f). While endocrine differentiation progresses as a wave throughout development (Johansson et al., 2007) *in vivo*, it is more synchronized in the hESC differentiation platform *in vitro* (Pagliuca et al., 2014; Rezania et al., 2014; Russ et al., 2015). At a timepoint directly preceding beta cell differentiation, we found that nearly 70% of hESC-derived cells were either *NGN3*+/*FEV*+ or *FEV*+ (Fig. 2.9f, blue and yellow bars). These data place the FEV+ population at a timepoint consistent with an endocrine progenitor population during human beta cell differentiation *in vitro*.

Endocrine dynamics over developmental time

Although we had captured comparatively fewer epithelial cells at E12.5 and E17.5 than at E14.5, we could still identify the Fev^{Hi} cells at both timepoints (Fig. 2.10a). To capture more epithelial cells and account for those that were missing from E12.5 and E17.5 version 1 (v1) runs, we re-performed an entirely new (version 2) set of single-cell RNA-sequencing experiments at E12.5, E14.5, and E17.5 after depletion of CD140a+ mesenchymal cells in order to enrich for epithelial cells (Fig. 2.10b,c). Given the high numbers of red blood cells at E17.5,

we ran two wells of E17.5 cells (replicates 1 and 2) to increase our capture of epithelial cells and then aggregated the datasets. We first analyzed the exocrine compartment and identified acinar, ductal, and proliferating populations of both at all timepoints (Fig. 2.10d and Supplementary Data 8-10). We then focused on the endocrine compartment, where we captured 584, 1,267, and 1,837 endocrine cells at E12.5, E14.5, and E17.5, respectively. We found similar gene expression topologies as in our v1 dataset but gained additional resolution with increased cell numbers and transcriptomic coverage (Fig. 2.10e and Supplementary Data 11-13).

To analyze how endocrine populations change over time, we merged all three v2 timepoints into one dataset using canonical correlation analysis (Butler et al., 2018). We correlated the v2 dataset to the v1 dataset and could identify all populations present in the v1 dataset (Fig. 2.10f and Supplementary Data 14). We also found additional populations, including a cluster characterized by decreased expression of *Fev* and increased expression of *Pdx1* and *Mafb*, genes with known roles in endocrine lineage decisions (Fig. 2.9h and Fig. 2.10g). This Pdx1+/Mafb+ population correlates most strongly with the Fev^{Hi}/Chgb+ population, as well as both the alpha and beta cell populations in the v1 dataset (Fig. 2.10f). We also found a second beta cell population characterized by increasing expression of *Ins1* and *Ins2* and lower expression of *Pdx1*, perhaps representing more mature beta cells (Fig. 2.10g). Indeed, this second beta cell group is almost entirely comprised of cells from the E17.5 timepoint (Fig. 2.9i). To examine how these populations shift over developmental time, we calculated the proportion of these populations at each timepoint (Fig. 2.9j). We found shifts in cell proportions that match those reported in literature, such as a high proportion of alpha cells early in development at E12.5 and increasing proportions of beta and delta cells at later timepoints (Johansson et al., 2007). The Ngn3+ population decreased over time, while the Fev+/Pax4+, Fev^{Hi}/Chgb+, and Pdx1+/Mafb+ populations peaked at E14.5, consistent with previous studies that reported peak *Ngn3* expression at approximately E14.5 and its subsequent downregulation as differentiation

into endocrine lineage ensues (Villasenor et al., 2008). At E17.5, we also found an increasing proportion of proliferating endocrine cells, presumably those responsible for the expansion of endocrine cell mass in later embryonic development (Bonner-Weir et al., 2016). These results from the larger v2 dataset confirm our initial findings from the v1 dataset and add additional resolution to the endocrine populations during pancreatic development.

Lineage decisions within the endocrine compartment

As the *in vivo* lineage tracing data had revealed that the Fev^{Hi} population is derived from the Ngn3+ population, we hypothesized that the Fev^{Hi} population could then function as a progenitor for the endocrine populations of the developing pancreas. We utilized a *Fev-Cre; ROSA26*^{mTmG} lineage tracing strategy to label *Fev*-expressing cells and their progeny. We found that 100% of alpha, beta, and delta cells, 90.1% of gamma cells, and 23.2% of epsilon cells were lineage-traced in E14.5 pancreas (Fig. 2.11a-e). These proportions of lineage labeling held true later in development (E17.5) and in adulthood (6 weeks) (Fig. 2.12 and 2.13). Epsilon cells are rare in the adult pancreas (Arnes et al., 2012) and still exhibited only partial lineage tracing in E17.5 pancreas (47.8% traced) (Fig. 2.12e). These results demonstrate that the majority of endocrine cells pass through a *Fev*-expressing stage during development.

We next combined this lineage tracing approach with single-cell RNA-sequencing to identify transcriptional regulators of endocrine differentiation. FACS sorting was used to enrich for *Fev*-expressing cells and their progeny (membrane-GFP+) from *Fev-Cre*; *ROSA26*^{mTmG} pancreata at E14.5 (Fig. 2.11f,g). All expected endocrine populations were identified in the resulting single-cell dataset (Fig. 2.11h,i). In addition, we found that *eGFP* reads mapped to all endocrine populations except the Ngn3+ population (Fig. 2.11i), further confirming that *Fev* expression turns on after *Ngn3*.

We next set out to model the lineage relationships among the endocrine cells and identify transcriptional regulators of differentiation. Pseudotime ordering identified a trajectory that began with Ngn3+ cells, transitioned into Fev+ cells, and then split into two main branches (Fig. 2.14a; see similar branching pattern in analysis of our first v1 dataset, Fig. 2.15a). The termini of the branches were populated by differentiated beta and alpha cells, suggesting that the branches represent a transition from a progenitor to fully differentiated hormone+ cell (Fig. 2.14a).

We next used Monocle's branched expression analysis modeling (BEAM) to identify the genes that distinguish the paths along the two branches to either alpha or beta cells. We found gene clusters that were upregulated along different segments of the pseudotime trajectory (Fig. 2.14b and Supplementary Data 15) and performed pathway analysis to identify pathways enriched at each stage of pseudotime (Fig. 2.15c and Supplementary Data 16). Genes upregulated at the beginning of pseudotime in gene cluster 2 included early markers of endocrine differentiation, such as *Sox4* and *Ngn3* (Fig. 2.14b). *Fev* was in gene cluster 6 and increased in both branches before ultimately decreasing in expression at the branch termini (Fig. 2.14b,c). Gene cluster 6 also included other genes expressed within the Fev^{Hi} population, including *Cldn4*, *Vim*, and *Chgb* (Fig. 2.14b,c and Fig. 2.15b). We found branch-specific clusters that included known markers of beta (*Ins1*) and alpha (*Gcg*) cells and known differentiation regulators of alpha (*Arx*, *Pou3f4, Irx1, Slc38a5,* and *Tmem27*) and beta (*Pdx1, Pak3,* and *Nkx6-1*) cells (Fig. 2.14c and Fig. 2.15b) (Pan and Wright, 2011; Petri et al., 2006; Stanescu et al., 2017). These clusters also contained genes that were enriched in either the alpha or beta branch but were expressed before acquisition of hormone expression (Fig. 2.15b). Within the alpha cell branch, *Peg10*, *Smarca1*, *Auts2*, and *Wnk3* increased in expression before upregulation of *Gcg* occurred (Fig. 2.15b). *Peg10* and *Auts2* have roles in differentiation (Dekel et al., 2006; Hishida et al., 2007) and migration (Hori et al., 2014), but a role in endocrine differentiation has not been described. As a regulator of chromatin states and an adult human

alpha cell marker (Muraro et al., 2016), *Smarca1* may be involved in the epigenetic regulation of alpha cell differentiation. Within the beta cell branch, *Gng12*, *Tssc4*, *Ece1, Tmem108*, *Wipi1* and *Papss2* increased in expression before upregulation of *Ins1* (Fig. 2.15b). To our knowledge, a role in endocrine lineage decisions have not been described for these beta branch-specific genes. We found a similar endocrine differentiation trajectory by an orthogonal method that uses force-directed layouts to visualize gene topologies and infer lineage relationships within single-cell data (Tusi et al., 2018; Weinreb et al., 2018) (Fig. 2.15d). We hypothesize that the genes identified by the analysis above may represent regulators of the differentiation of an endocrine progenitor to a fully differentiated hormone-expressing cell.

To validate our pseudotime results, we performed ISH for markers that defined each branch of the trajectory. First, we confirmed the expression of *Peg10* and *Gng12* within the Fev^{Hi} population (Fig. 2.14d,e, indigo and teal gradient arrows), validating the expression of these genes in a stage before hormone acquisition. We also validated the enrichment of *Peg10* and *Gng12* in alpha and beta cells, respectively (Fig. 2.14f,g, solid indigo and teal arrows). First, 95.8% of beta cells expressed *Gng12* (n=46 cells, 6 pancreata), while 30.5% expressed *Peg10* (n=71 cells, 7 pancreata) (Fig. 2.14f and Fig. 2.16a). Additionally, 100% of alpha cells expressed *Peg10* (n=31 cells, 6 pancreata), while only 5.4% expressed *Gng12* (n=32 cells, 4 pancreata) (Fig. 2.14g and Fig. 2.16b). The lineage relationships generated by pseudotime ordering, combined with the validation *in vivo*, lead us to hypothesize that the Fev+/Peg10+ cells are fated towards an alpha cell identity and Fev+/Gng12+ cells towards a beta cell identity (Fig. 2.14h). These results suggest that lineage allocation of endocrine progenitors towards alpha or beta cell fates may occur after the onset of *Fev* expression.

Discussion

The mesenchyme is critical for epithelial specification and proliferation throughout pancreatic development (Ahlgren et al., 1997; Golosow and Grobstein, 1962; Landsman et al., 2011), yet the individual cell types responsible for these processes remain unidentified. Our single-cell dataset has enabled the identification of multiple mesenchymal subpopulations and gene candidates for regulating epithelial-mesenchymal interactions. Secreted factors, such as mesothelial-derived Fgf9, may play a similar role in the pancreas as in the lung, where it regulates mesenchymal cell proliferation and vascular formation (Yin et al., 2011). Additionally, secretion of Wnt antagonists by cluster 5 may regulate Wnt signaling in the developing pancreas, influencing processes such as epithelial specification, expansion, and exocrine development (Murtaugh, 2008). Future work can focus on uncovering the functions of these individual mesenchymal populations in development, physiology, and pathology of the pancreas.

With the cell types of the mesenchyme now enumerated and their markers identified, we can begin to elucidate the maturation and lineage relationships of the pancreatic mesenchymal compartment. Our time course data have provided evidence of maturation within the mesothelial population. Genes such as *Pitx2*, *kallikren 13* (*Klk13)* and 8 (*Klk8*), were differentially expressed in younger, E12.5 mesothelial cells. *Pitx2* regulates differentiation in multiple systems (Cao et al., 2013; Hernandez-Torres et al., 2017; Shang et al., 2008), and the kallikren family are serine proteases involved in extracellular matrix and adhesive molecule degradation (Kapadia et al., 2004). Expression of these genes suggests that the E12.5 mesothelial population may be primed for migration and differentiation. In contrast, the E17.5 mesothelial population expressed genes related to barrier or immune function, such as *dermokine* (*Dmkn*) (Hasegawa et al., 2013; Huang et al., 2017), *bone marrow stromal antigen 2* (*Bst2*), and *retinoic acid receptor responder 2* (*Rarres2*) (Ernst and Sinal, 2010). These results suggest stage-dependent roles for the mesothelium throughout development.

The different roles for the mesothelium across time are also evident from our pseudotime analysis, which proposes that the mesothelium serves as a progenitor for other mesenchymal cell types during development. The mesothelium is a critical mesenchymal progenitor population in other organs, such as the heart, intestine, lung, and liver (Wilm et al., 2005). Our data suggest that mesothelial progenitor activity occurs at E12.5 or earlier during pancreatic development, consistent with other organ systems (Bin Zhou et al., 2011; Que et al., 2008; Winters and Bader, 2013). Indeed, a recent study identified that parietal mesothelial cells can function as progenitor cells prior to pancreatic specification (Angelo and Tremblay, 2018). The transcriptomic information obtained by this study will allow the development of tools to target individual populations within the mesenchyme and perform lineage tracing, ablation, and expression studies *in vivo*. Furthermore, this developmental dataset can be compared to mesenchymal population dynamics during adult disease progression, where aberrant recapitulation of developmental pathways can lead to disease states in the pancreas (Jensen et al., 2005; Rhim and Stanger, 2010). Thus, this dataset is a broad resource for the implementation of future studies in pancreatic mesenchymal biology.

Within the epithelial compartment, our identification of a Fev^{Hi} endocrine progenitor population provides increased resolution of endocrine differentiation. The relative timing of expression of canonical endocrine lineage genes can now be mapped onto these additional differentiation stages. Several lines of evidence suggest that the gene *Fev* may be a direct target of Ngn3: *Fev* is the transcription factor most strongly expressed in Ngn3+ endocrine progenitors (Miyatsuka et al., 2009), and *Ngn3* knockout embryos do not express *Fev* in the developing pancreas (Ohta et al., 2011). Known target genes of *Ngn3*, such as *Pax4* (Collombat et al., 2003) and *Runx1t1i* (Benitez et al., 2014), are expressed by the early-stage Fev+/Pax4+ population. Additionally, *Pax6* is upregulated within the Fev^{Hi} population. Although *Chga* and *Chgb* are often utilized as markers of differentiated endocrine lineages, we found that both are expressed in the Fev^{Hi} population prior to hormone acquisition. This result is consistent with

previous work that identified Chga+, hormone- cells in rodent pancreatic development (Butler et al., 2016). The Fev^{Hi} cell stage likely represents the cell stage during endocrine differentiation preceding specialized hormone production and may now serve as a cellular landmark for understanding endocrine lineage gene expression dynamics.

The gene *Fev* has been previously studied mainly in serotonergic neurons, where it is a master transcriptional regulator required for cellular differentiation, maturation, and serotonin synthesis (Spencer and Deneris, 2017). Fev switches transcriptional targets from differentiation genes during development to maturation genes postnatally in serotonergic neurons (Wyler et al., 2016). In an insulinoma cell line, Fev directly binds to the regulatory regions of serotonergic genes, such as *Tph1*, *Tph2*, *Ddc*, *Slc18a2*, and *Slc6a4*, as well as the *Ins1* promoter itself (Ohta et al., 2011). Future ChIP-seq studies of embryonic pancreas will globally identify direct targets of Fev and Fev-regulated transcriptional networks in developing endocrine cells.

Using genetic lineage tracing *in vivo*, we have demonstrated that the majority of endocrine cells in the developing pancreas transit through a *Fev*-expressing stage, and that Fev-lineage cells contribute not only to embryonic, but also to adult pancreatic endocrine cells. The fraction of epsilon cells not derived from a Fev-lineage may represent the subset of Ghrl+ cells previously reported to give rise to the ductal and exocrine lineages (Arnes et al., 2012). As all adult gamma cells are Fev-lineage labeled, the small subset of gamma cells not lineage traced during pancreatic development may represent those that do not persist in the adult pancreas.

Further highlighting the relevance of Fev^{Hi} progenitors during pancreatic development, our pseudotime analysis revealed that *Fev*-expressing cells may be pre-specified towards an alpha or beta cell fate. As expected, we found expression of *Ins1* and *Gcg* at the termini of the beta and alpha branches, and upregulation of *Pdx1* and *Arx*, which are known regulators of endocrine cell fate decisions, earlier in pseudotime. In addition, our pseudotime analysis identified genes enriched along the alpha or beta branch and expressed prior to upregulation of

hormones. These genes warrant further study as potential regulators of the acquisition of alpha or beta cell identity.

For the eventual application of this knowledge to human therapeutics, the findings in the murine model must be validated in human pancreatic development. Our staining of human fetal pancreas identified the analogous FEV^{Hi} population, consistent with our findings in murine pancreata. Directed differentiation of hESCs towards endocrine cell fates will provide a platform for modeling and manipulating the putative lineage regulators found in this study. Indeed, we have identified a FEV+ population within hESC-derived endocrine progenitor cells. Deeper knowledge of these lineage decisions may substantially improve directed differentiation efforts to efficiently generate functional beta cells for cellular replacement therapy for people with diabetes. This study highlights the power of combining single-cell transcriptomic information with *in vivo* lineage tracing to reconstruct developmental trajectories within cellular compartments. Discovery of populations and their lineage relationships will promote identification of the mechanisms that drive lineage decisions and commitment.

Materials and Methods

Mice

All mouse procedures were approved by the University of California, San Francisco (UCSF) Institutional Animal Care and Use Committee (IACUC). Mice were housed in a 12-hour light-dark cycle in a controlled temperature climate. Noon of the day of vaginal plug was considered embryonic day 0.5.

Timed-pregnant Swiss Webster mice were obtained from Charles River Laboratories. *Ngn3-Cre* (Schonhoff et al., 2004) (a gift from Dr. Matthias Hebrok), *Fev-Cre* (Scott et al., 2005) (The Jackson Laboratory 012712), and *ROSA26*^{mTmG} (Muzumdar et al., 2007) (the Jackson Laboratory 007676) mice were maintained in a C57BL/6J background. The *Cre* transgene was genotyped using the following primers: GGGCGGCATGGTGCAAGTT and CGGTGCTAACCAGCGTTTTC.

Human tissue procurement and isolation

Human fetal pancreata were harvested from post-mortem fetuses at 23 weeks of gestation with permission from the ethical committee of the University of California, San Francisco (UCSF). Tissue was fixed in 4% paraformaldehyde overnight at 4°C. After three washes in 1X PBS, tissue was either cryopreserved in 30% sucrose solution at 4°C overnight and embedded in OCT, or placed in 40% ethanol then 70% ethanol before paraffin embedding. 8 um sections were cut on the cryostat or microtome. *In situ* hybridization and immunofluorescence were then performed as described below.

Adult human islets were isolated from cadaveric donor tissue by the UCSF Islet Production Core with permission from the UCSF ethical committee. Consented cadaver donor pancreata were provided by the nationally-recognized organization UNOS via local organ procurement agencies. The identifiers were maintained at the source only, and the investigators received de-identified specimens.

Informed consent was obtained for all human (fetal and adult) tissue collection, and protocols were approved by the Human Research Protection Program Committee on Human Research of the University of California, San Francisco (UCSF).

Embryonic stem cell culture and differentiation

The human embryonic stem cell (hESC) line HUES8 was obtained from Harvard University and used for the generation of hESC-derived beta-like cells (BLCs). Pluripotent HUES8 cells were maintained as spherical clusters in suspension in mTeSR-1 (StemCell Technologies) in 500mL spinner flasks (Corning, VWR) on a magnetic stir plate (Dura-Mag) within a 37°C incubator at 5% CO_2, 100% humidity, and a rotation rate of 70 rpm. Cells were screened for mycoplasma contamination using the MycoProbe Mycoplasma Detection Kit (R&D Systems), according to the manufacturer's instructions.

hESC-derived endocrine progenitor cells were generated as previously described (Pagliuca et al., 2014). In brief, HUES8 cells were seeded into a spinner flask at a concentration of 8 x 10^5 cells/mL in mTeSR-1 media with 10μM Rock inhibitor Y27632 (StemCell Technologies) to allow formation of spherical clusters. Differentiation was initiated 72 hours later. Differentiation was achieved in a step-wise fashion using the following growth factors and/or small molecules: definitive endoderm (Stage 1) (1 day of 100 ng/mL Activin A (R&D Systems) and 14 μg/mL of CHIR99021 (Stemgent); 2 days of 100 ng/mL Activin A); gut tube endoderm (Stage 2) (3 days of 50 ng/mL KGF (Peprotech)); early pancreatic progenitors (Stage 3) (1 day of 200 nM LDN193189 (Fisher Scientific), 50 ng/mL KGF, 0.25 μM Sant-1 (Sigma), 2 μM Retinoic Acid (Sigma), 500 nM PdbU (EMD Biosciences); 1 day of 50 ng/mL KGF, 0.25 μM Sant-1, 2 μM Retinoic Acid, 500 nM PdbU); later pancreatic progenitors (Stage 4) (5 days of 50 ng/mL KGF, 0.25 μM Sant-1, 0.1 μM Retinoic Acid); endocrine progenitors (Stage 5) (4 days of 0.25 μM Sant-1, 0.1 μM Retinoic Acid, 1 μM XXI (EMD Millipore), 10 μM Alk5i (Axxora), 1 μM T3 (EMD Biosciences), 20 ng/mL Betacellulin (Fisher Scientific); 3 days of 25 nM Retinoic Acid,

1 μM XXI, 10 μM Alk5i, 1 μM T3, 20 ng/mL Betacellulin); BLCs (Stage 6) (6 days of 10 μM Alk5i; 1 μM T3). Successful differentiation was assessed at Stages 1, 3, 4, 5, and 6 via immunofluorescence or FACS for stage-specific marker genes.

To measure the expression of *FEV* at various stages of human endocrine differentiation, aliquots of clusters were removed from the flask and analyzed at several timepoints: after 5 days in Stage 5 ("mid-stage endocrine progenitors"), after 7 days in Stage 5 ("late-stage endocrine progenitors"), and after 5 days at the BLC stage. As a comparator, pluripotent, undifferentiated hESCs in mTeSR-1, as well as human adult islets, were also analyzed for *FEV* expression.

Immunofluorescence

Embryonic mouse pancreata were dissected in cold 1X PBS and fixed in zinc-buffered formalin (Anatech LTD) at room temperature (RT) for 30-90 minutes or overnight at 4°C. After three washes in 1X PBS, tissue was processed for either cryopreservation or paraffin embedding. Cryopreserved pancreata were placed in 30% sucrose solution at 4°C overnight before embedding in OCT. Paraffin-embedded pancreata were placed in 40% ethanol and 70% ethanol before paraffin tissue processing. 8 um sections were cut on the cryostat or microtome. For immunofluorescence on paraffin sections, slides were baked at 55°C for 30 minutes, deparaffinized in xylene, and rehydrated in decreasing concentrations of ethanol. Heat-mediated antigen retrieval was performed using Antigen Retrieval Citra Solution (Biogenex Laboratories). Tissue sections were blocked in 5% normal donkey serum (NDS; Rockland Immunochemicals) and Mouse-on-Mouse IgG blocking reagent (Vector Laboratories) when appropriate in 0.2% Triton X-100 in PBS (PBT) for 1 hour and then stained overnight at 4°C using the following primary antibodies: Acta2 (1:200, Abcam ab21027), Cav1 (1:200, Abcam ab2910), Chromogranin A (1:100, Abcam ab15160), E-cadherin (1:200, BD Transduction Lab 610182), Glucagon (1:100, Abcam ab82270), Insulin (1:50, DAKO A0564), Vimentin (1:200,

Abcam ab92547), and Wt1 (1:100, Abcam ab89901). All antibodies have been validated by the manufacturer. The next day, sections were washed three times in 0.1% Tween 20 in 1X PBS and then incubated with species-specific Alexa Fluor 488-, 594-, or 647-conjugated secondary antibodies (1:500, Jackson ImmunoResearch) and DAPI in 5% NDS in 0.2% PBT for 1 hour at RT. Sections were washed three times in 0.1% Tween 20 in 1X PBS, rinsed in 1X PBS, and then mounted in Fluoromount-G mounting medium (SouthernBiotech). Slides were stored at 4°C.

For immunofluorescence on cryosections, slides were removed from -80°C storage and allowed to reach RT. Sections were rinsed in 1X PBS three times and permeabilized in 0.5% PBT for 10 minutes at RT. Tissue sections were blocked in 5% NDS and, if needed, Mouse-on-Mouse IgG blocking reagent in 0.1% PBT for 1 hour and then stained overnight at 4°C using the following primary antibodies: Epcam (1:200, BD Transduction Lab 552370), Glucagon (1:2000, Millipore 4031-01F), Insulin (1:250, DAKO A0564), Somatostatin (1:500, Santa Cruz Biotechnology sc-7819, Ghrelin (1:1500, Santa Cruz Biotechnology sc-10368), Pancreatic Polypeptide (PP; 1:250, Abcam ab77192), and Vimentin (1:200, Abcam ab92547). All antibodies have been validated by manufacturer. Sections were washed the next day three times in 1X PBS and then incubated with species-specific Alexa Fluor 488-, 555-, 594-, or 647-conjugated secondary antibodies and DAPI in 5% NDS in 0.1% PBT for 1 hour at RT. Sections were washed three times in 1X PBS and mounted in Fluoromount-G mounting medium. Slides were stored at 4°C.

Images were captured on a Zeiss Apotome Widefield microscope with optical sectioning capabilities or Leica confocal laser scanning SP8 microscope. Maximum intensity z-projections were then prepared using ImageJ, where brightness, contrast, and pseudo-coloring adjustments were applied equally across all images in a given series.

In situ hybridization

In situ hybridization was performed on 8 um sections using RNAscope technology (Advanced Cell Diagnostics) (Wang et al., 2012) according to the manufacturer's instructions. *In situ* probes against mouse *Ngn3* (422409-C2)*, Fev* (413241-C3)*, Isl1* (451931)*, Ins1* (414661-C4), *Gcg* (400601)*, Sst* (404631-C3)*, Ghrl* (415301-C2)*, Ppy* (482701)*, Peg10* (512921-C4)*, Gng12* (462521-C2)*, Nnat* (432631-C2)*, Barx1* (414681)*, Pitx2* (412841-C2)*, Stmn2* (498391-C3)*, Msln* (443241) and human *NGN3* (505791-C4), *FEV* (471421-C3), and *ISL1* (478591-C2) were used in combination with the RNAscope Multiplex Fluorescent Reagent Kit v2 for target detection. Following signal amplification of the target probes, sections were washed in 1X PBS three times and blocked in 5% NDS in 0.1% PBT for 1 hour at RT. Tissue sections were then stained with primary and secondary antibodies as described above in the "immunofluorescence" section.

For *in situ* hybridization of hESC-derived clusters, cells were fixed with 4% PFA for 15 minutes at RT, washed with PBS, and cryoprotected in 30% sucrose overnight. The next day, clusters were embedded in a small sphere of 1.5% low-melting temperature agarose; these were again cryoprotected in 30% sucrose overnight. The following day, the agarose spheres were soaked in OCT and frozen in a dry ice bath. *In situ* hybridization was then performed on 8 um sections using human *NGN3*, *FEV*, and *ISL1* RNAscope probes.

Quantification of cell proportions

Quantification of pancreata was performed by manual counting using ImageJ software. Cell populations present at less than 1% in *Ngn3*-lineage traced E14.5 replicates were deemed artifact and excluded from further analysis.

Quantitative RT-PCR

hESCs from various stages of directed differentiation were collected and RNA extracted with the RNeasy Mini Kit (Qiagen). Reverse transcription was performed with the Clontech RT-PCR kit. RT-PCR was run on a 7900HT Fast Real-Time PCR instrument (Applied Biosystems) with Taqman probes for *FEV* (assay ID: Hs00232733_m1) and *GAPDH* (assay ID: Hs02758991_g1) in triplicate. Data were normalized to *GAPDH*. Error bars represent standard deviation.

Dissociation and FACS of embryonic pancreas

Embryonic mouse pancreata were dissected and placed in 1X PBS on ice, then dissociated into single cells using TrypLE Express dissociation reagent (Thermo Fisher) at 37°C with pipet trituration at 5-minute intervals during incubation. For v1 datasets, E12.5 pancreata were dissociated for 10 minutes, E14.5 pancreata for 15 minutes, and E17.5 pancreata for 30 minutes. For batch 1, we pooled 14 E14.5 pancreata from one litter. For batch 2, which was collected on a different day, we pooled tissue from each timepoint separately: 18 E12.5 pancreata from two litters, 11 E14.5 pancreata from one litter, and 8 E17.5 pancreata from one litter. Dissociations were neutralized with FACS buffer (10% FBS + 2mM EDTA in phenol-red free HBSS). Dissociated cells were passed through a 30 um cell strainer and stained with Sytox live/dead stain (Thermo Fisher). Stained cells were washed twice in FACS buffer and then sorted using a BD FACS Aria II. After size selection to remove doublets, all live cells were collected.

For version 2 10X datasets, we pooled tissue from each time point separately, each performed on a different day: 14 E12.5 pancreata from one litter, 13 E14.5 pancreata from one litter, and 13 E17.5 pancreata from one litter. For the E14.5 *Fev-Cre; ROSA26*^{mTmG} 10X sample, we pooled 3 pancreata from one litter. Dissociations were performed as described above. Cells undergoing a CD140a negative selection were stained with CD140a-APC (1:50; eBiosciences,

cat. 17-1401-81; validated by manufacturer). Stained cells were washed twice in FACS buffer and then sorted using a BD FACS Aria II. After size selection to remove doublets, all live CD140a- cells were collected. For the E14.5 *Fev-Cre; mTmG* pancreata, live GFP+ cells and GFP-/TdTomato+ cells were collected. All 4,000 GFP+ (*Fev*-lineage traced) cells were loaded onto the 10X Genomics platform, supplemented with an additional 21,000 TdTomato+/GFP- (non-lineage traced).

Single-cell capture and sequencing

To capture individual cells, we utilized the Chromium Single Cell 3' Reagent Version 1 Kit (10X Genomics) (Zheng et al., 2017). For batch 1, 12,800 cells from E14.5 pancreata were loaded into one well of the 10X chip, while for batch 2, 18,000 cells per time point were each loaded into their own respective wells to produce Gel Bead-in-Emulsions (GEMs). GEMs underwent reverse transcription to barcode RNA before cleanup and cDNA amplification. Libraries were prepared with the Chromium Single Cell 3' Reagent Version 1 Kit. Each sample was sequenced on 2 (Batch 1) or 1 (Batch 2) lanes of the HiSeq2500 (Illumina) in Rapid Run Mode with paired-end sequencing parameters: Read1, 98 cycles; Index1, 14 cycles; Index2, 8 cycles; and Read2, 10 cycles.

The CD140a-depleted E12.5, E14.5, and E17.5 datasets and *Fev-Cre; ROSA26*^{mTmG} dataset in Figures 5 and 7 were generated with Chromium Single Cell 3' Reagent Version 2 kits (10X Genomics). 27,000 cells were loaded onto their respective wells and underwent the same processing as the Version 1 kits, according to manufacturer instructions for Version 2 kits. Libraries were sequenced on the NovaSeq (Illumina) with the same sequencing parameters as above.

Single-cell analysis

For the v1 datasets, we utilized CellRanger v1.1.0 software for v1 datasets and v2.1.0 for v2 datasets with default settings for de-multiplexing, aligning reads to the mouse genome (10X Genomics pre-built mm10 reference genome) with STAR (Dobin et al., 2012) and counting unique molecular identifiers (UMIs) to build transcriptomic profiles of individual cells. For the v1 datasets, gene-barcode matrices were analyzed with the R package Seurat v1.4, using the online tutorial as a guide (Satija et al., 2015). We first performed a filtering step, retaining only the cells that expressed a minimum of 200 genes and only the genes that were expressed in at least 3 cells. A large number of cells did not meet this threshold in the E17.5 time point and were determined to be red blood cells by the high expression of hemoglobin genes. Variable genes were determined by mean-variance relationship to identify highly-expressed and variable genes with the Seurat function MeanVarPlot with default settings. UMI counts were log-normalized, and linear regression was performed with RegressOut to account for differences in the number of UMIs between cells. Principal component analysis (PCA) was then utilized to determine sources of variability in the dataset with PCAfast. Significant PCs were determined based on the Scree plot and utilized for Seurat's graph-based clustering algorithm (function FindClusters) with default parameters, except for the resolution parameter. To vary cluster numbers, the resolution parameter in FindClusters was adjusted from 0.6 – 3.0, and resulting clusters analyzed as follows. Clusters were visualized with t-distributed stochastic neighbor embedding (t-SNE) with Seurat's RunTSNE function with default settings (Maaten and Hinton, 2008). Differentially-expressed genes were determined with the FindAllMarkers function, which uses a bimodal likelihood ratio test (McDavid et al., 2012). We confirmed differential gene expression analysis with the Wilcoxon rank sum test and MAST (Finak et al., 2015) utilizing Seurat v2's FindMarkers function with default settings. These tests calculate adjusted p-values for multiple comparisons. To determine the final number of clusters, clusters were required to have at least 9 significantly ($p < 0.05$) differentially expressed genes with a 2-fold difference in

expression in comparison to all other clusters. Clusters were manually curated for differential gene expression, and those that did not meet this threshold were manually merged with the nearest cluster based on the phylogenetic tree from Seurat's BuildClusterTree. In some cases, clusters met the 9-gene threshold but appeared to have very similar differentially-expressed genes to another cluster. This is likely a result of the comparison of individual clusters against all other clusters in determining differentially-expressed genes. In these cases, a pairwise comparison between the two clusters was performed and the same 9-gene threshold applied. An exception to the 9-gene threshold was made to annotate the proliferating population in early stages of the cell cycle within the E14.5 mesenchymal analysis (Fig. 2.4, cluster 8). Additionally, cluster 10 in the E14.5 mesenchymal dataset did not meet the 9-gene threshold. Rather, clusters 1-9 had distinct transcriptomic signatures (with at least 9 differentially expressed genes) that distinguished them from cluster 10. Lists of at least 2-fold differentially-expressed genes for individual analyses are provided in Supplementary Data 2.

For v2 datasets, Seurat v2.2 and v2.3 was utilizing to perform the analysis. Cells with less than 200 genes and genes expressed in fewer than 3 cells were removed, as above. UMI counts were normalized with NormalizeData using default settings. Variable genes were determined with FindVariableGenes, using the following cut-offs suggested by the online tutorial (x.low.cutoff = 0.0125, x.high.cutoff = 3, y.cutoff = 0.5). Data was scaled and UMI counts regressed out with the ScaleData function. Principal component analysis was performed with RunPCA, and significant PCs determined based on the Scree plot. t-SNE analysis and clustering was performed as described above for the v1 datasets. For the E12.5 exocrine dataset, the ductal population did not meet the 9-gene threshold. All other populations within this dataset could be distinguished from the ductal population by at least 9-differentially expressed genes, therefore we still annotated this cluster. Some of the clusters depicted for the *Fev-Cre; ROSA26*^{mTmG} dataset do not meet the 9-gene threshold. We chose to visualize these clusters in order to better illustrate their placement along the pseudotime trajectory.

Custom genome build

The custom genome for alignment of reads to eGFP and TdTomato sequences from the mTmG mouse line was created according to instructions provided by 10X Genomics reference support (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/advanced/references). eGFP and TdTomato sequences were concatenated to the mm10-2.1.0 reference genome (FASTA file) provided by 10X Genomics. eGFP and TdTomato annotations were then concatenated to the mm10 annotations (GTF file) provided by 10X Genomics. The cellranger mkref command was then utilized with the genome and annotations with eGFP and TdTomato, as described in the above link.

Pathway analysis

Pathway analysis and calculation of associated p-values were performed using the ConsensusPathDB over-representation analysis for pathway-based sets category (http://cpdb.molgen.mpg.de) (Kamburov et al., 2008).

Aggregating E17.5 v2 datasets

E17.5 technical replicates from the v2 dataset were aggregated with Cellranger v2.1, utilizing the aggr function with default settings. The aggregated dataset was used for analysis and merging with the E12.5 and E14.5 v2 datasets.

Sub-clustering and merging datasets

Sub-clustering was performed by isolating clusters of interest with the Seurat function SubsetData and reanalyzing as described above (identification of variable gene, regression, and determination of significant PCs). Cells were classified as epithelial based on the expression of *E-cadherin* (*Cdh1*) and other known epithelial population markers. Cells that were *Cdh1-, Vim*+, and *collagen3a1* (*Col3a1*)+ were classified as mesenchymal. Multiple batches were merged

with the MergeSeurat function. The merged dataset was reanalyzed as above, with batch included as a latent variable in the RegressOut function. The v1 E14.5 batch 1 and batch 2 clusters were robust to the sampling differences between batches as evidenced by the contribution of cells from both batches to each cluster (Fig. 2.3b). We find high correlation of cell type proportion between batches in all populations except the exocrine compartment (acinar and ductal) (Fig. 2.3c), possibly due to technical challenges of pancreatic dissociation. Within each cluster, batch 1 cells correlated most highly with those of batch 2 contained in the same cluster, indicating proper cluster calling with the merged datasets (Fig. 2.3d).

For v2 datasets (E12.5, E14.5 and E17.5), multiple canonical correlation analysis (multiCCA) from Seurat v2.3 was utilized to merge the epithelial datasets (Butler et al., 2018). The top 1,000 most highly variable genes that were variable in at least 2 datasets were used for the alignment, as recommended in the Seurat tutorial. The shared correlation strength of each CC was measured with Seurat's MetageneBicorPlot, and those before the drop-off were used for alignment, analogous to the Scree plot in choosing significant PCs. We then aligned the datasets with AlignSubspace and ran an integrated t-SNE and clustering analysis, as outlined in the Seurat tutorial. Clusters were required to have 9 significantly differentially expressed genes as described above. Clusters with similar differentially expressed genes were verified with pairwise comparisons to the most related clusters (based on BuildClusterTree) and merged if they did not meet the pairwise 9-gene threshold. The Beta 2 cluster in the v2 endocrine merged timecourse data met the 9-gene threshold for 2 out of the 3 differential expression tests (Bimodal likelihood ratio and Wilcoxon rank sum tests), but had only 8 differentially expressed genes for the MAST test.

Doublets were identified based on co-expression of two mutually exclusive genes, such as both mesenchymal and epithelial genes, and removed from further analysis. In the v2 datasets, rare cells (4 cells in E12.5 and 13 cells in E14.5 endocrine datasets) with high levels of hemoglobin gene expression were removed from the analysis.

Downsampling analysis

To determine if the sequencing depth was sufficient for calling clusters, downsampling analysis was performed for the v1 E14.5 batch 1 dataset. Reads were randomly downsampled from the 10x Cellranger bam file output to a specified percentage, then grouped based on UMI to generate a count profile for each cell. The number of genes with greater than 0 counts was then calculated. UMI downsampling was performed with the SampleUMI function. A new Seurat object was created with the downsampled matrix and reanalyzed as above.

The number of UMIs/cell was downsampled from an average of 4,600 UMIs/cell in the full dataset to 200 UMIs/cell, and the median number of genes/cell and clustering robustness was then calculated. Clustering robustness was determined as the percentage of cells within the same cluster, with clusters required to maintain at least 9 genes with a 2-fold change in expression in comparison to all other clusters. Within this dataset, robust clustering was maintained all the way down to 500 UMIs/cell, when the percentage of cells in the same cluster began to climb, indicating collapsing of individual clusters. Both of these downsampling analyses indicate that sufficient sequencing depth was reached.

Pseudotemporal ordering

We utilized Monocle 2.6.4 (Qiu et al., 2017b) to order cells in pseudotime based on their transcriptomic similarity. For v1 timecourse datasets, batch-corrected values and variable genes from Seurat analysis were used as input, utilizing the gaussianff expressionFamily, and clusters were projected onto the minimum spanning tree after ordering.

For the *Fev*-lineage traced dataset, UMI counts and variable genes from the Seurat analysis were used as input, utilizing the negBinom expressionFamily. To find genes differentially-expressed across the branch point in the trajectory, we used monocle's internal BEAM analysis and selected genes with an FDR cutoff of 0.001. Gene expression patterns were plotted with plot_genes_branched_heatmap and plot_multiple_branches_pseudotime.

Code availability

Seurat and monocle R objects used for analysis are available, along with scripts, at Figshare, DOI: 10.6084/m9.figshare.c.4158458. https://figshare.com/collections/_/4158458

Scripts are available at https://github.com/sneddonucsf/2018-Developmental-single-cell-RNA-sequencing.

Data availability

RNA sequences for the single-cell RNA-sequencing analyses reported in this paper have been deposited in the GEO database under accession code GSE101099.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE101099

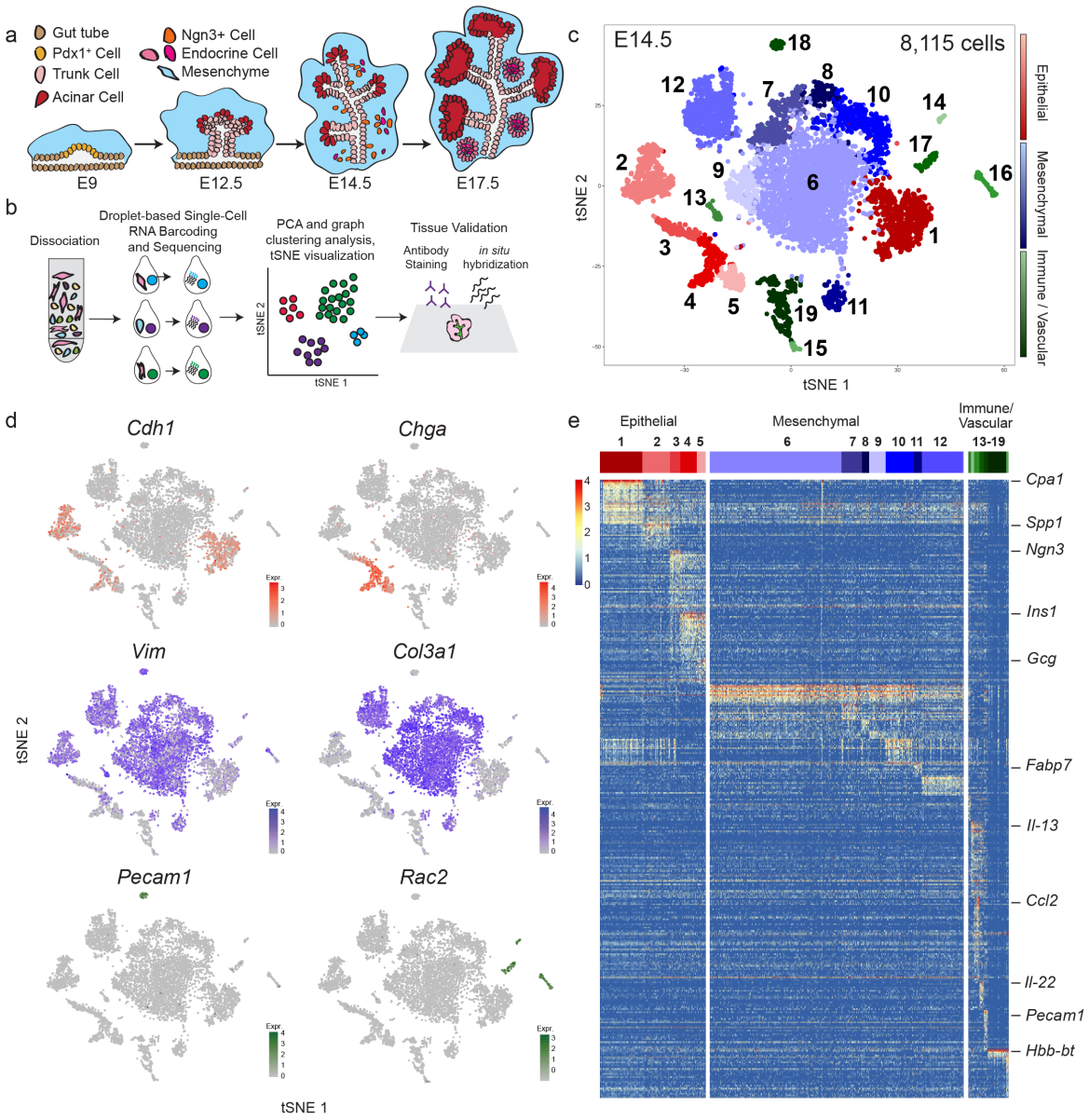
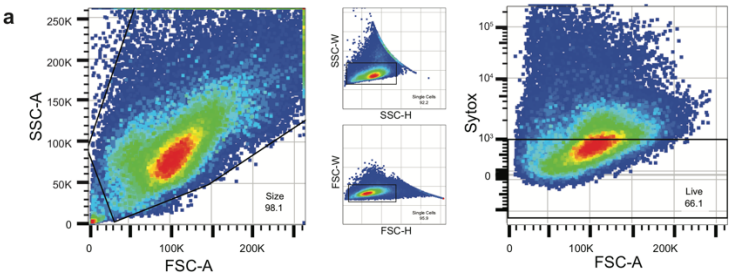


Figure 2.1. Single-cell sequencing identifies broad patterns of cellular heterogeneity in E14.5 murine pancreas. (a) Overview of murine pancreatic development. (b) Schematic of experimental approach. (c) t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization of populations from pooled E14.5 mouse pancreata (n=14). Each dot represents the transcriptome of a single cell, color-coded according to its cellular identity (epithelial, mesenchymal, or immune/vascular). Each cell compartment contains multiple sub-populations, represented by varying degrees of color shading. (d) Established marker genes identify epithelial cells (Cdh1+), endocrine cells (Chga+), mesenchymal cells (Vim+ and Col3a1+), endothelial cells (Pecam1+), and immune cells (Rac2+). (e) Heatmap depicting greater than 2-fold differentially-expressed genes in each cluster compared to all other clusters. Cells are represented in columns, and genes in rows. Specific genes used to annotate clusters are indicated to the right of the heatmap.

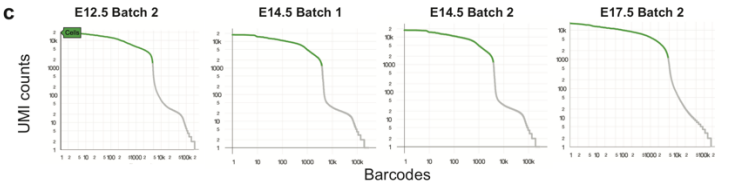
a



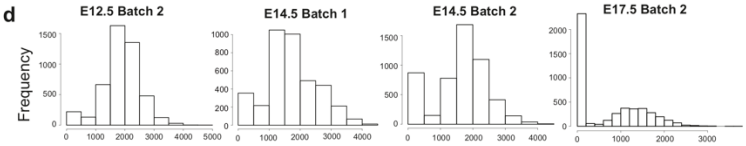
b

	E12.5 Batch 2	E14.5 Batch 1	E14.5 Batch 2	E17.5 Batch 2
Cells Captured	4,631	3,849	5,179	4,635
Median reads/cell	28,016	61,681	31,875	25,324
Median UMIs/cell	4,339	3,084	4,388	3,419
Median genes/cell	1,912	1,621	1,731	194
Valid Barcodes	75.70%	94.20%	92.50%	75.90%
Mapped Reads to transcriptome	65.40%	64.10%	66.70%	66.20%
Sequencing Saturation	58.60%	87.70%	71.20%	63.70%
Fraction Reads in Cells	90.10%	90.80%	93.00%	92.30%

c



d



e

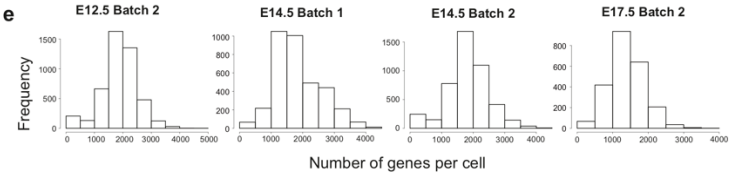


Figure 2.2. Quality Control for Single-cell RNA-sequencing Runs. (a) Representative FACS plot of single, live cells sorted from dissociated Swiss Webster embryonic pancreata and used for single-cell sequencing. (b) Quality control statistics for all single-cell sequencing runs prepared with the Chromium Single Cell 3' Reagent Version 1 Kit. The "valid barcodes" metric indicates the percentage of cells with barcodes that match a known barcode contained on a bead. "Mapped reads to transcriptome" refers to the percentage of reads that confidently map to a unique gene in the reference transcriptome. "Fraction Reads in Cells" is the percentage of reads that contain a cell-associated barcode. (c) Cellranger cell calls based on the number of UMIs. The dropoff indicates the threshold for the number of UMIs required for a barcode to be considered a cell. (d) Histogram of the number of genes per cell in all single-cell runs pre-filtering steps. (e) Histogram of the number of genes per cell in all single-cell runs post-filtering steps. E17.5 Batch 2 contained a large number of red blood cells, which expressed fewer than 200 genes, resulting in their removal during minimum gene threshold filtering (see Methods).

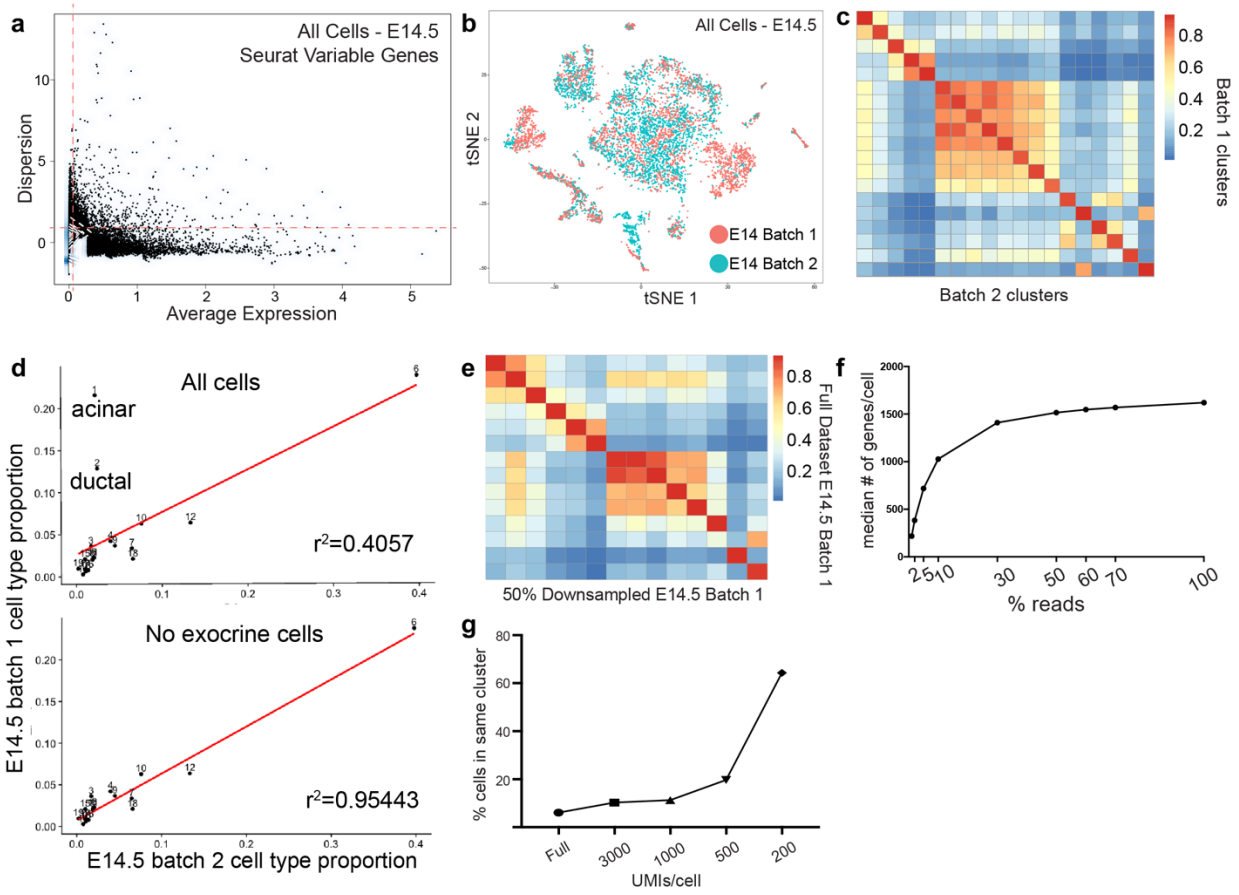


Figure 2.3. Single-cell RNA-sequencing Batch Information from E14.5 Pancreata. (a) Selection of variable genes in the E14.5 dataset (all cells) by Seurat's MeanVarPlot function. (b) t-SNE visualization of merged E14.5 batches, color-coded by batch. Batch 1 and 2 contribute to all clusters, reflecting a successful batch correction. (c) Pearson's correlation of E14.5 batch 1 cells with E14.5 batch 2 cells within each cluster based on average expression of variable genes. Batch 1 cells correlate most highly with batch 2 cells within the same cluster, indicating proper merging of the two batches. (d) Cell type proportions in E14.5 batch 1 and 2 with exocrine (acinar and ductal) clusters included (top panel) and excluded (bottom panel). All cell types except the exocrine compartment show high correlation between the two batches. (e) Pearson's correlation between clusters from the E14.5 batch 1 full dataset and those from the E14.5 batch 1 dataset downsampled to 50% of the reads, based on average expression of shared variable genes. (f) Maintenance of the number of median genes/cell after random downsampling of reads, indicating sufficient sequencing depth. (g) Maintenance of cluster structure after random downsampling of UMIs is reflected by the similar percentage of cells found within the same cluster with fewer UMIs.

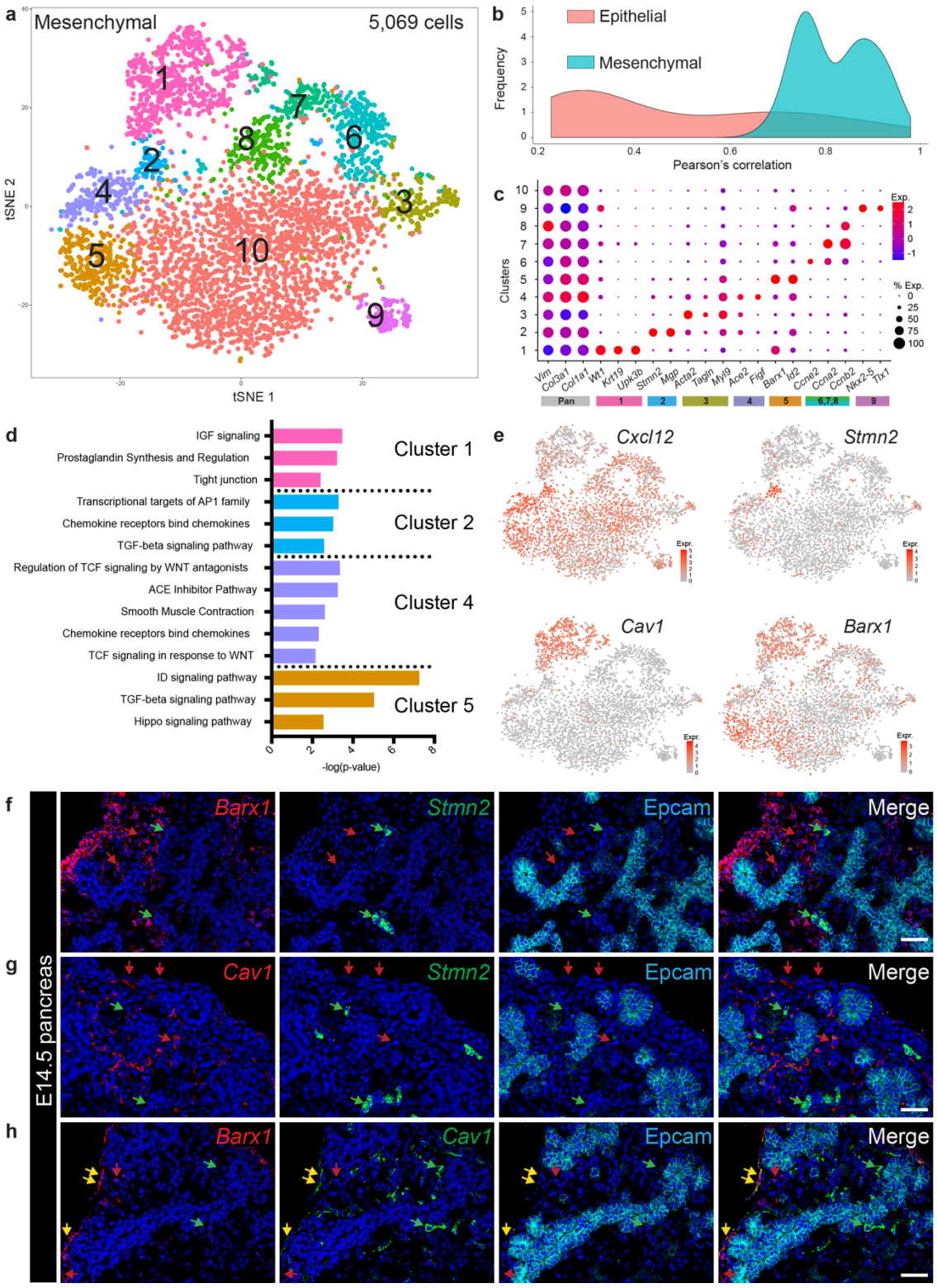


Figure 2.4. Identification of multiple uncharacterized mesenchymal populations. (a) t-SNE visualization of subclustered E14.5 mesenchymal clusters (from n=14 pancreata). (b) Density plot depicting Pearson's correlation values (depicted in heatmap in Fig. 2.5b) within the epithelial and mesenchymal populations based on average gene expression in each cluster. (c)

Dot plot of top differentially-expressed markers of each mesenchymal population. Bars are color-coded by cluster identity in (a). The grey bar represents pan-mesenchymal markers. The size of each dot represents the proportion of cells within a given population that expresses the gene; the intensity of color indicates the average level of expression. (d) Pathway analysis of genes greater than 2-fold differentially-expressed by cells in clusters 1, 2, 4, and 5. (e) Expression of genes marking clusters 1 (*Cav1*), 2 (*Stmn2*), 4 (*Cxcl12*), and 5 (*Barx1*) in all E14.5 mesenchymal cells. Color intensity indicates level of expression. (f-h) Multiplexed fluorescent ISH combined with Epcam IF validates clusters 2 and 5 (e) and cluster 1 (f-g). Epcam marks pancreatic epithelium. In (f), Barx1+ cells (red arrows, cluster 5) are distinct from Stmn2+ cells (green arrows, cluster 2), validating the single-cell data. In (g), Cav1+ cells (red arrows, cluster 1) are distinct from Stmn2+ cells (green arrows, cluster 2). In (h), Barx1+ cells that do not express *Cav1* (red arrows) represent cluster 5, whereas Barx1+/Cav1+ cells (yellow arrows) represent cluster 1. Cav1+ cells that do not express *Barx1* are also identified (green arrows), likely representing endothelial cells (Frank, 2003). Scale bar represents 50 um in f-h.

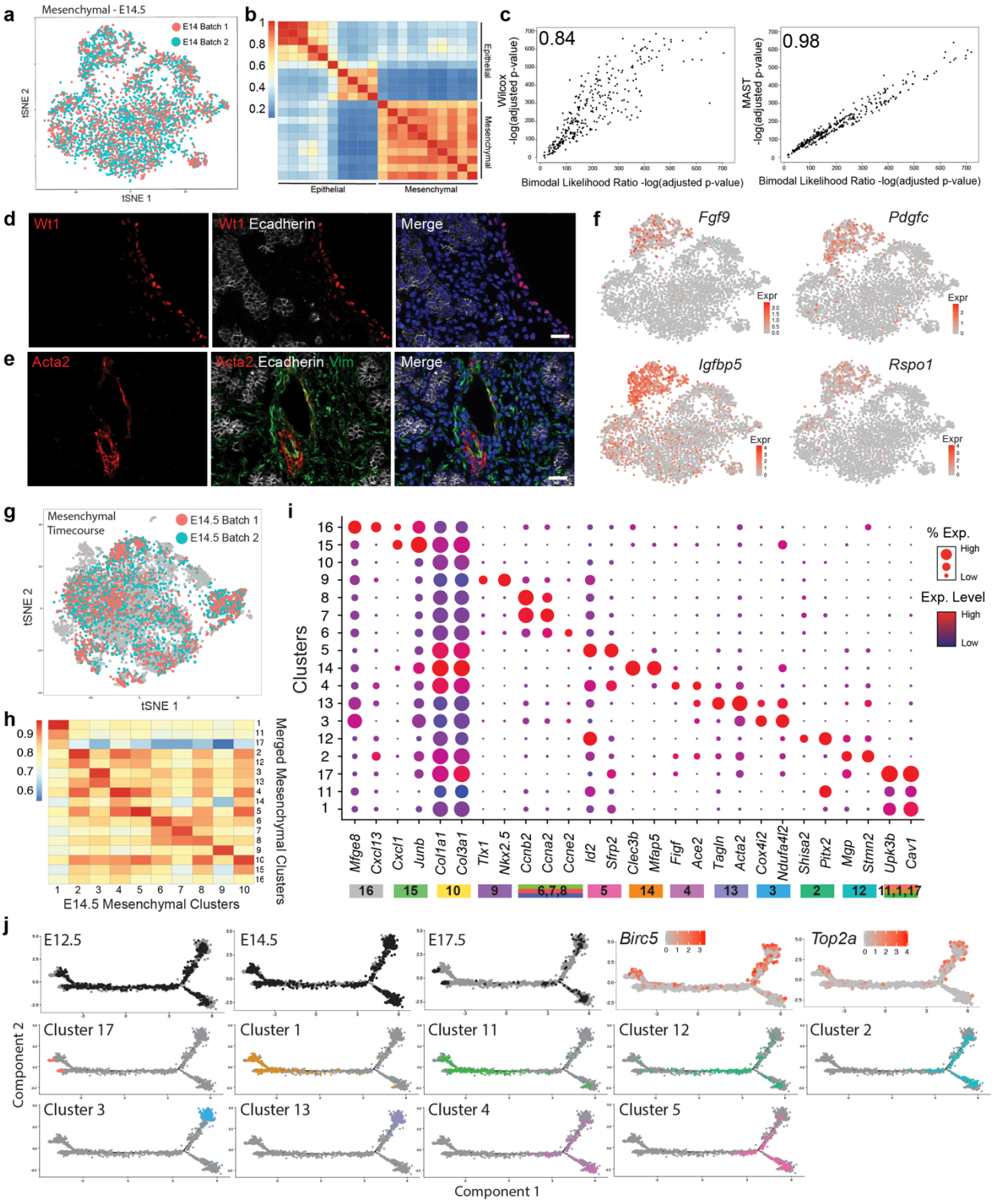


Figure 2.5. Transcriptomic Signatures and Lineage Dynamics among Mesenchymal Populations. (a) t-SNE visualization of E14.5 biological replicates, colored by batch, demonstrating effectiveness of batch correction. (b) Pearson's correlation of E14.5 epithelial and mesenchymal clusters based on average expression of variable genes. (c) Comparison of bimodal likelihood ratio test adjusted p-values to adjusted p-values calculated by either MAST (left panel) or Wilcox rank sum (right panel) tests for all greater than 2-fold differentially-

expressed genes. Pearson's correlation value is shown in top left corner. (d,e) IF validation of (d) mesothelium (Wt1+) and (e) vascular smooth muscle (Acta2+) cells in E14.5 pancreata. Ecadherin marks epithelium, and Vim marks mesenchyme. Scale bar: 50 um. (f) Expression of secreted factors within the mesothelium. Color indicates level of expression. (g) t-SNE visualization of merged mesenchymal timecourse dataset. E14.5 biological replicates are colored, serving as a measure of batch correction effectiveness within the merged mesenchymal timecourse dataset. Grey dots represent both E12.5 and E17.5 cells. (h) Correlation of E14.5 mesenchymal populations with merged (E12.5, E14.5 and E17.5) mesenchymal clusters based on average expression of the variable genes from all datasets. Merged populations were matched with E14.5 (Fig. 2.4) by highest correlation and assigned the same cluster identity (cluster 1-10). Remaining merged clusters were assigned cluster identities 11-17. (i) Dot plot of differentially-expressed genes from each merged mesenchymal cluster. Colored bars correspond to t-SNE in Fig. 2.14a (j) Contribution of cells from each timepoint is mapped onto pseudotime plots. Expression of proliferation markers, *Birc5 and Top2a*, in the pseudotime trajectory. Color indicates level of expression. Contribution of cells from each timepoint is broken down by individual cluster and mapped onto pseudotime plots. Colors correspond to cell clusters in Fig. 2.4a,e.

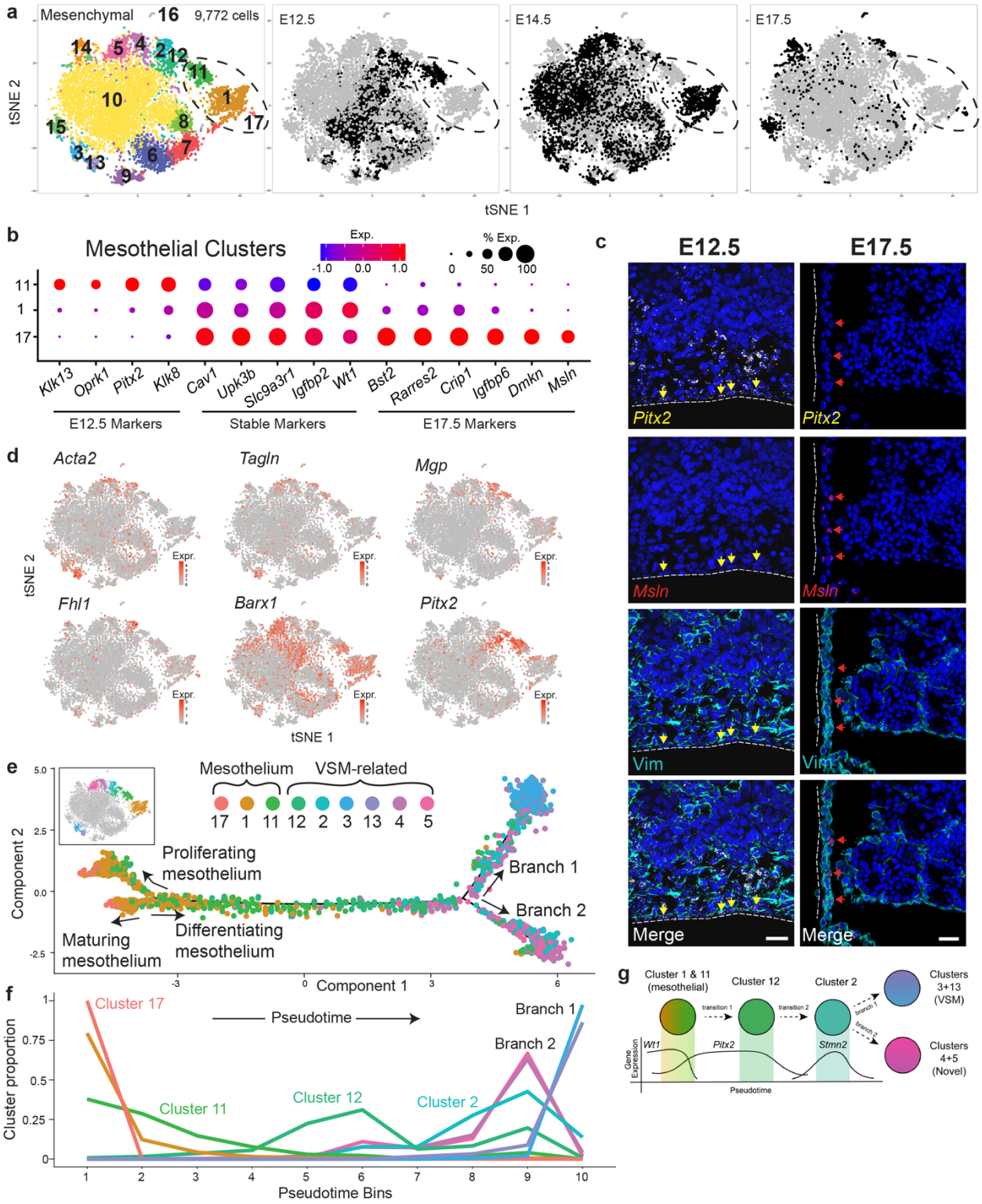


Figure 2.6. Mesothelial cells are dynamic over developmental time and are predicted to give rise to vascular smooth muscle populations. (a) t-SNE visualization of merged mesenchymal clusters from E12.5 (n=18 pancreata), E14.5 (n=14 pancreata for batch 1; n=11 for batch 2), and E17.5 (n=8 pancreata) tissue. Mesenchymal clusters were identified at each timepoint, subclustered, merged together, and reanalyzed. Cells are colored by cluster or

timepoint. Dotted circle highlights timepoint-segregated mesothelial clusters. (b) Dot plot of top differentially-expressed genes in timepoint-specific mesothelial clusters (clusters 1, 11, and 17). Size of the dot represents proportion of the population that expresses each specified marker. Color indicates level of expression. (c) ISH for *Pitx2* and *Msln* in E12.5 and E17.5 pancreata. *Pitx2* expression was detected in E12.5 but not E17.5 mesothelium, whereas *Msln* was detected in E17.5 but not E12.5 mesothelium. Vimentin (Vim) IF staining depicts pancreatic mesenchyme. Dotted line indicates tissue boundary. Yellow arrows identify Pitx2+ mesothelial cells. Red arrows identify Msln+ mesothelial cells. Scale bar represents 50 um. (d) Expression levels of VSM-related genes in merged mesenchymal clusters. Color intensity indicates level of expression. (e) Pseudotime ordering of mesothelial and VSM-related merged mesenchymal clusters. Colors correspond to t-SNE in (a). All clusters are individually plotted in Fig. 2.5j. (f) Cluster proportions over pseudotime. Pseudotime was binned into 10 groups and the proportion of each cluster within that bin of pseudotime was calculated. (g) Model of lineage relationships among mesothelial, vascular smooth muscle, and VSM-related mesenchymal populations based on pseudotime ordering in (e).

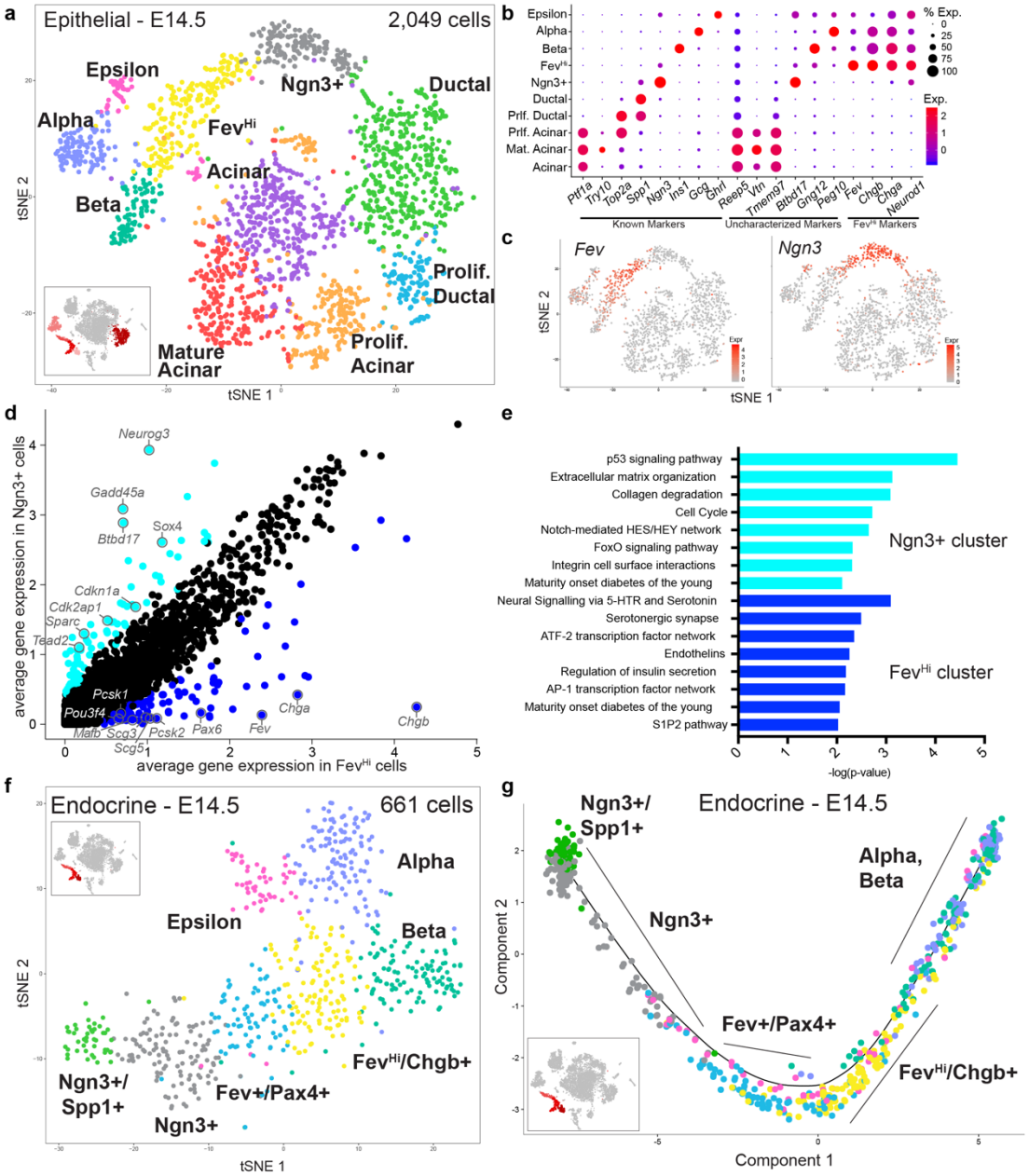


Figure 2.7: Identification of epithelial cell populations in E14.5 mouse pancreas. (a) t-SNE visualization of epithelial groups only, as defined in Figure 1. (b) Dot plot depicting known and uncharacterized markers of epithelial populations, as well as markers specific to the Fev^{Hi} population. Size of the dot represents proportion of the population that expresses each specified marker. Color indicates level of expression. (c) Expression of *Fev* and *Ngn3* within epithelial cells. Color indicates level of expression. (d) Gene expression comparison between the Ngn3+ and Fev^{Hi} population. Genes greater than 2-fold differentially-expressed are highlighted in dark blue (higher in Fev^{Hi} cells) or light blue (higher in Ngn3+ cells). (e) Pathway analysis of genes greater than 2-fold differentially-expressed in Ngn3+ and Fev^{Hi} populations (f) t-SNE visualization of the 661 cells of the endocrine lineage (Ngn3+, Fev^{Hi}, alpha, beta, and epsilon populations). (g) Pseudotime ordering of Ngn3+, Fev+/Pax4+, Fev^{Hi}, alpha, and beta cell populations place Fev+ cells between Ngn3+ and hormone+ populations.

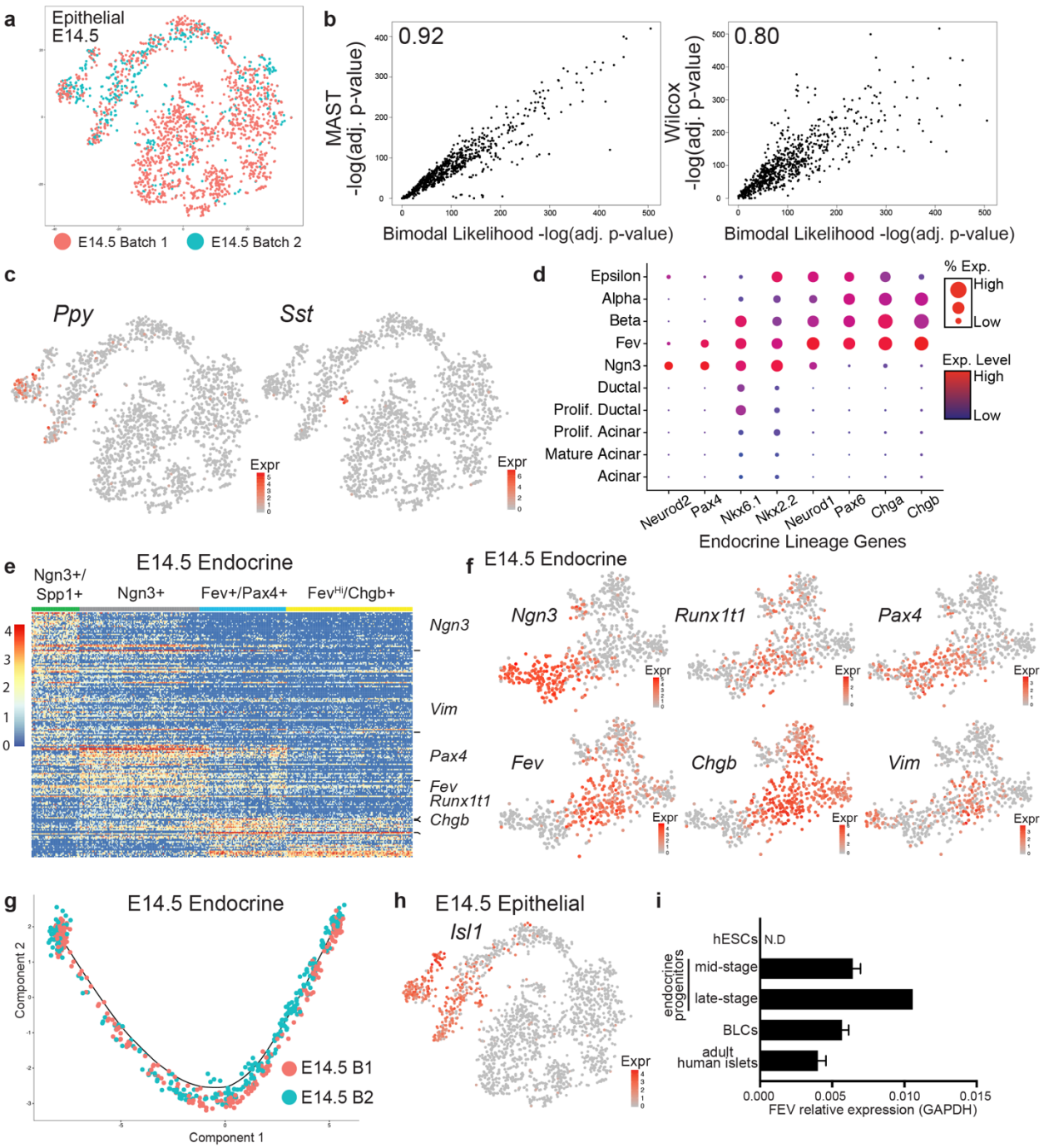


Figure 2.8. Identification of Known and Novel Epithelial Cell Populations in E14.5 Pancreas. (a) t-SNE visualization of E14.5 epithelial batches, colored by batch. Significant overlap and most importantly, clusters that include cells from both batches, reflects successful batch correction. (b) Comparison of bimodal likelihood ratio test adjusted p-values to adjusted p-values calculated by either MAST (left panel) or Wilcox rank sum (right panel) tests for all greater than 2-fold differentially-expressed genes. Pearson's correlation value is shown in top left corner. (c) Expression maps of *Ppy* and *Sst* hormones within E14.5 epithelial dataset. (d) Dot plot of endocrine lineage genes across the epithelial populations. (e) Heatmap depicting genes over 2-fold differentially-expressed in Ngn3+ and Fev+ populations. Differentially expressed genes were determined from the endocrine dataset depicted in Fig. 2.7f and only

Ngn3+ and Fev+ populations were shown in the heatmap. (f) Expression of selected markers of early- and late-Fev+ populations in all endocrine cell lineages. (g) Pseudotime ordering of Ngn3+, Fev+/Pax4+, Fev^{Hi}, alpha, and beta cell populations, colored by batch. (h) Expression of *Islet1* (*Isl1*) in E14.5 epithelial cells is largely confined to hormone+ populations. (i) Quantification of *FEV* expression by quantitative RT-PCR in pluripotent hESCs, mid- and late-stage endocrine progenitor cells, beta-like cells (BLCs), and adult human islets. *FEV* expression is normalized to *GAPDH*. Error bars represent standard deviation. N.D = not detected.

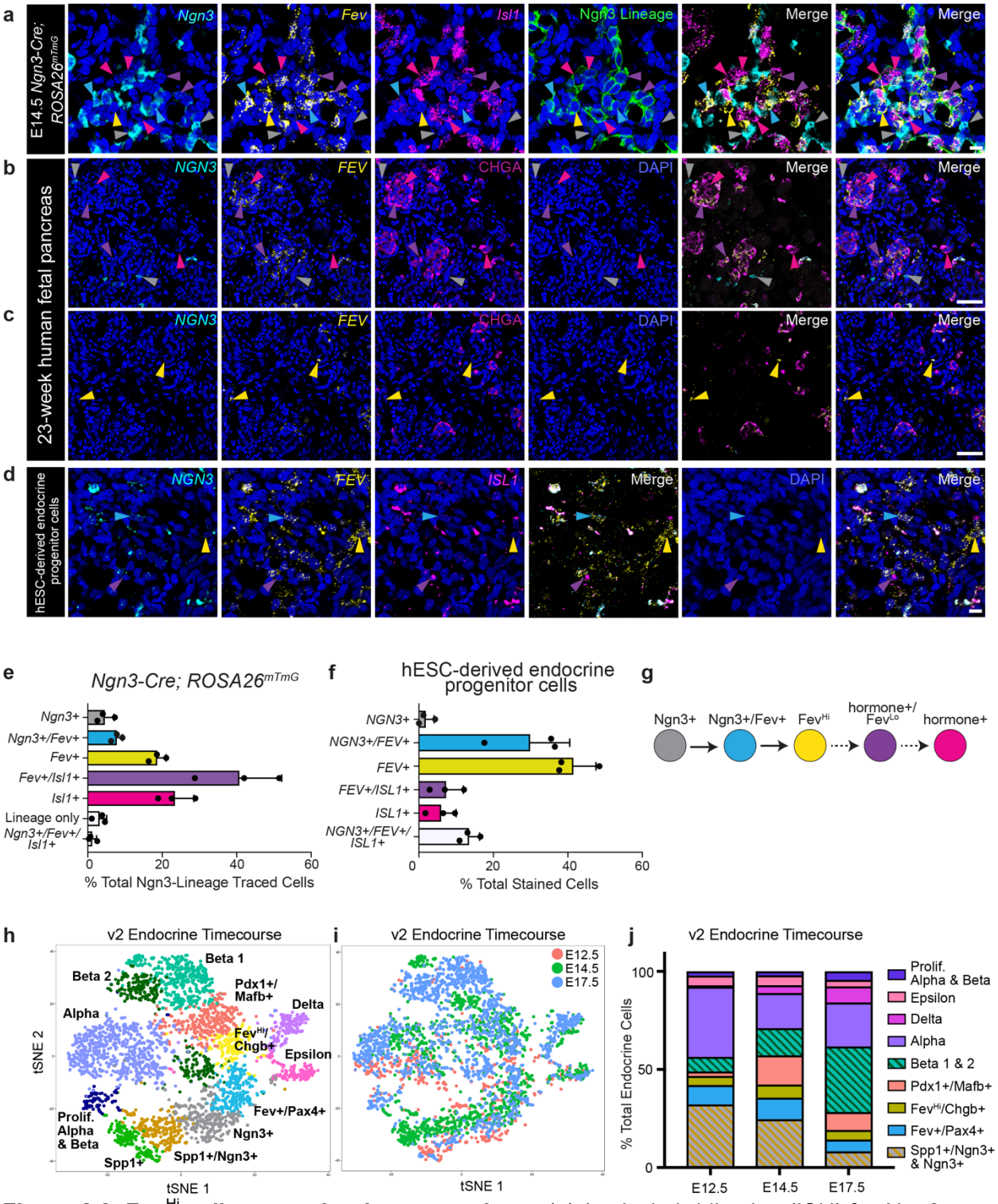
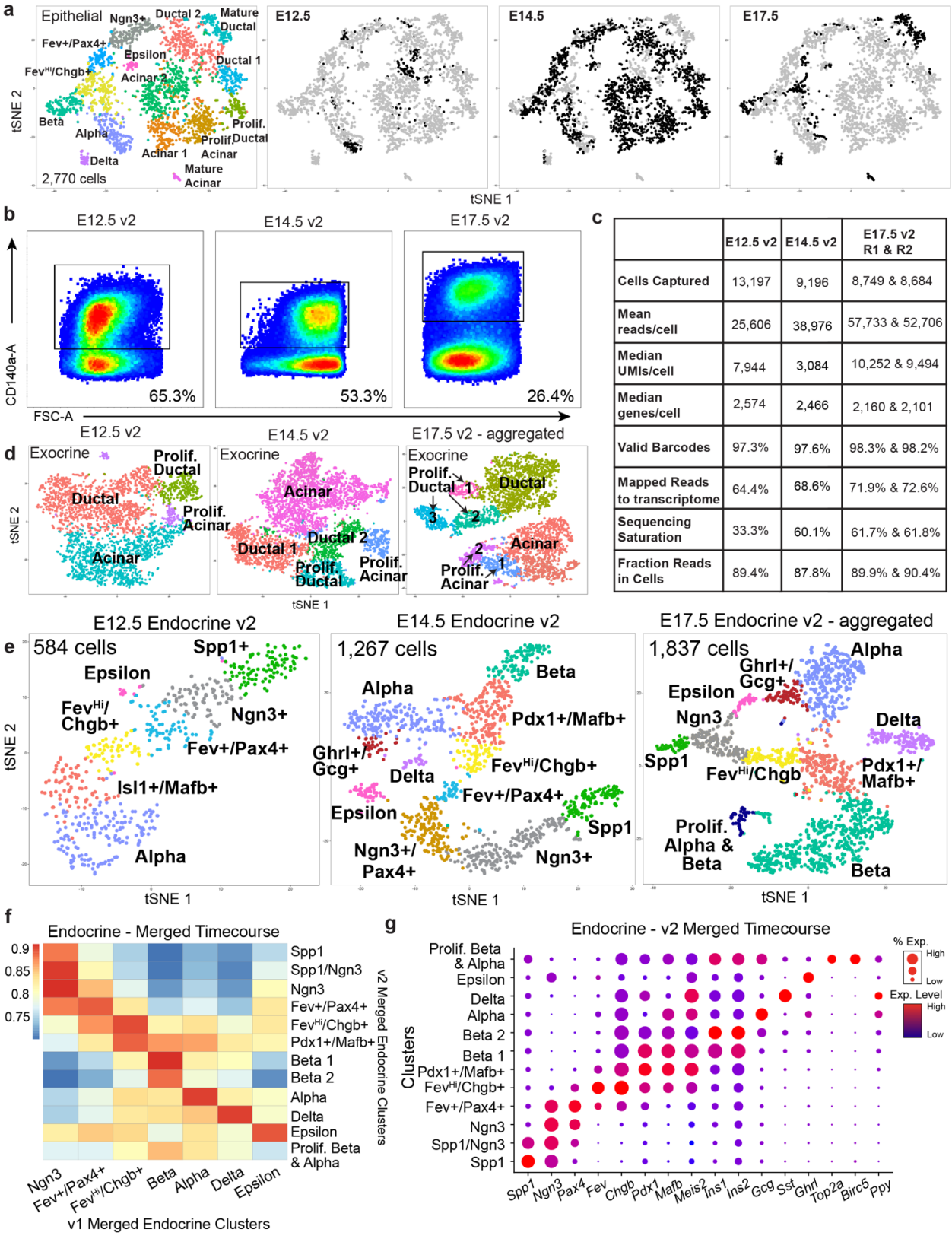


Figure 2.9. Fev^{Hi} cells are endocrine progenitors. (a) *In situ* hybridization (ISH) for *Ngn3*, *Fev*, and *Isl1* in lineage-traced *Ngn3-Cre; Rosa26^{mTmG}* E14.5 pancreata where *Ngn3*-lineage traced cells are mGFP+. Grey arrowheads identify *Ngn3*+ cells, presumably not yet Ngn3-lineage labeled due to the transient nature of *Ngn3* expression and the delay of Cre-mediated recombination that permits expression of mGFP. Blue arrowheads identify *Ngn3*+/*Fev*+ cells

that are Ngn3-lineage traced. Yellow arrowheads identify *Ngn3*-lineage traced cells that are *Fev*+ but do not express *Ngn3* or *Isl1*. Purple arrowheads identify *Fev*+/*Isl1*+ cells that are Ngn3-lineage traced. Magenta arrowheads identify *Isl1*+ cells that are Ngn3-lineage-traced. (b-c) Dual ISH/immunofluorescence (IF) for *NGN3* and *FEV* mRNA and CHGA protein in human fetal pancreas at 23 weeks of gestation (n=1 pancreas). Grey arrowheads identify *NGN3*+ cells. Yellow arrowheads identify *FEV*+ cells. Purple arrowheads identify *FEV*+/CHGA+ cells. Magenta arrowheads identify CHGA+ cells. (d) Multiplexed fluorescent ISH for *NGN3*, *FEV*, and *ISL1* mRNA in hESC-derived endocrine progenitor cells. Blue arrowheads identify *NGN3*+/*FEV*+ cells. Yellow arrowheads identify *FEV*+ cells. Purple arrowheads identify *FEV*+/*ISL1*+ cells. (e) Quantification of each population detected in *Ngn3*-lineage traced pancreata as a percentage of *Ngn3*-lineage traced cells (n=464 cells, 6 pancreata). Data are represented as mean + standard deviation (SD). (f) Quantification of each population detected in hESC-derived progenitor cells as a percentage of total stained cells (n=418 cells, 3 clusters representing technical replicates from one hESC differentiation). Data are represented as mean + SD (g) Proposed model for the derivation of Fev^{Hi} endocrine cells from Ngn3+ cells, and their differentiation into hormone+/Fev^{Lo} endocrine cells. Colors of arrowheads and bars in a-f correspond to cell identity in g. (a and d) Scale bar: 10 um. (b and c) Scale bar: 20 um. (h) t-SNE visualization of v2 merged endocrine timecourse (E12.5, E14.5, aggregated E17.5). Clusters are annotated based on correlation with v1 dataset or top differentially-expressed genes. (i) Timepoint labels for v2 merged endocrine timecourse data. t-SNE is the same as Fig. 2.9h. (j) Cell type proportions at each timepoint, calculated from the clusters depicted in Fig. 2.9h.



	E12.5 v2	E14.5 v2	E17.5 v2 R1 & R2
Cells Captured	13,197	9,196	8,749 & 8,684
Mean reads/cell	25,606	38,976	57,733 & 52,706
Median UMIs/cell	7,944	3,084	10,252 & 9,494
Median genes/cell	2,574	2,466	2,160 & 2,101
Valid Barcodes	97.3%	97.6%	98.3% & 98.2%
Mapped Reads to transcriptome	64.4%	68.6%	71.9% & 72.6%
Sequencing Saturation	33.3%	60.1%	61.7% & 61.8%
Fraction Reads in Cells	89.4%	87.8%	89.9% & 90.4%

Figure 2.10. Epithelial Populations over Developmental Time. (a) t-SNE visualization of merged version 1 epithelial clusters from E12.5 (n=18 pancreata), E14.5 (n=14 pancreata for

batch 1; n=11 for batch 2), and E17.5 (n=8 pancreata). All panels depict the same t-SNE plot. In the far-left panel, cluster identity is denoted by different colors. Throughout figure, cells are color-coded by cluster identity. In the three remaining panels, cells from the indicated timepoint are represented by black dots; all cells from the other timepoints are gray. (b) FACS plots depicting CD140a negative selection from E12.5 (n=14), E14.5 (n=13), and E17.5 (n=13) pancreata. CD140a-negative cells were used for single-cell sequencing. (c) Quality control statistics for 10X Chromium version 2 single-cell RNA-sequencing runs. These datasets are referred to as v2 datasets. Two technical replicates of E17.5 cells were run from the same pancreata on two separate wells on the 10X Chromium machine. The two E17.5 runs were aggregated and analyzed as one dataset. (d) Individual t-SNE plots of v2 E12.5, E14.5, and E17.5 (aggregated) exocrine dataset. Clusters are annotated based on gene expression. (e) Individual t-SNE plots of v2 E12.5, E14.5, and E17.5 (aggregated) endocrine dataset. Clusters are annotated based on correlation with v1 datasets and differentially expressed genes. (f) Pearson's correlation among clusters from v1 merged endocrine timecourse and v2 merged endocrine timecourse. (g) Dot plot of top differentially expressed genes for clusters in the v2 merged endocrine dataset. Clusters correspond to those depicted in t-SNE in Fig. 2.9h.

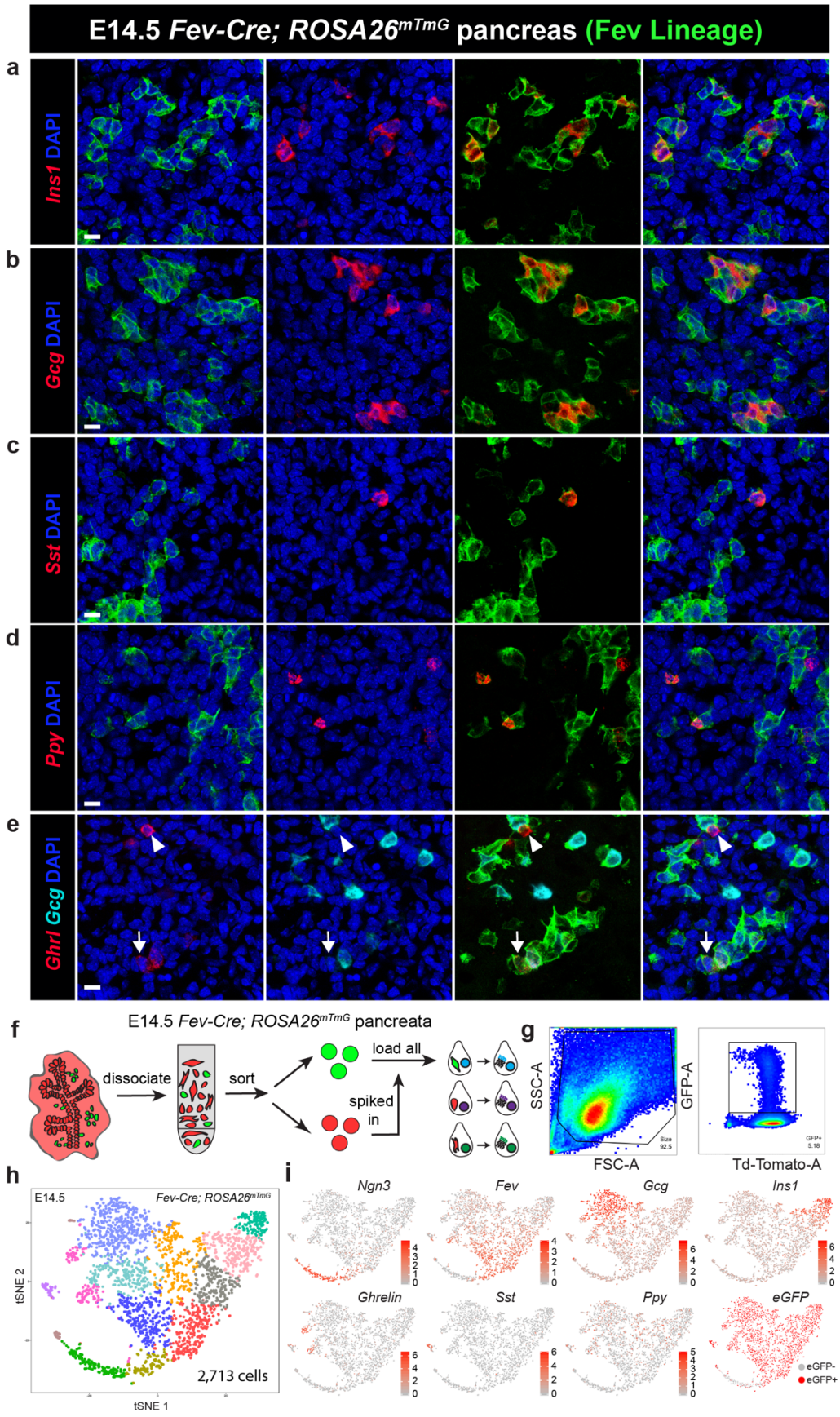

E14.5 Fev-Cre; ROSA26mTmG pancreas (Fev Lineage)
a
Ins1 DAPI
b
Gcg DAPI
c
Sst DAPI
d
Ppy DAPI
e
Ghrl Gcg DAPI
f
E14.5 Fev-Cre; ROSA26mTmG pancreata
dissociate
sort
load all
spiked in
g
SSC-A
FSC-A
GFP-A
Td-Tomato-A
h
E14.5
Fev-Cre; ROSA26mTmG
tSNE 2
tSNE 1
2,713 cells
i
Ngn3
Fev
Gcg
Ins1
Ghrelin
Sst
Ppy
eGFP
eGFP-
eGFP+

Figure 2.11. Differentiated hormone+ endocrine cells transit through a *Fev*-expressing stage during pancreatic development. (a-e) Dual IF (for membrane GFP) and fluorescent ISH for hormones in *Fev-Cre; ROSA26*^{mTmG} lineage traced animals at E14.5. n=46 cells of 4 pancreata for *Ins1* (100% labeled-lineage); n=103 of 4 pancreata cells for *Gcg* (100% lineage-labeled); n=6 cells of 2 pancreata for *Sst* (100% lineage-labeled); n=26 cells of 2 pancreata for *Ghrl/Gcg* (23.2% lineage-labeled); n=71 cells of 8 pancreata for *Ppy* (90.1% lineage-labeled). Scale bar represents 10um. (f) Schematic of E14.5 *Fev-Cre; ROSA26*^{mTmG} FACS sorting and single-cell RNA-sequencing. (g) Representative FACS plots of sorted single, live GFP+ and TdTomato+/GFP- cells from dissociated pancreata used for single-cell sequencing. (h) t-SNE visualization of endocrine cells in *Fev*-lineage traced E14.5 mouse pancreata (n=3). (i) Expression of major markers of endocrine cell types. Color indicates level of expression, except for the *eGFP* plot, which indicates presence or absence of *eGFP* counts.

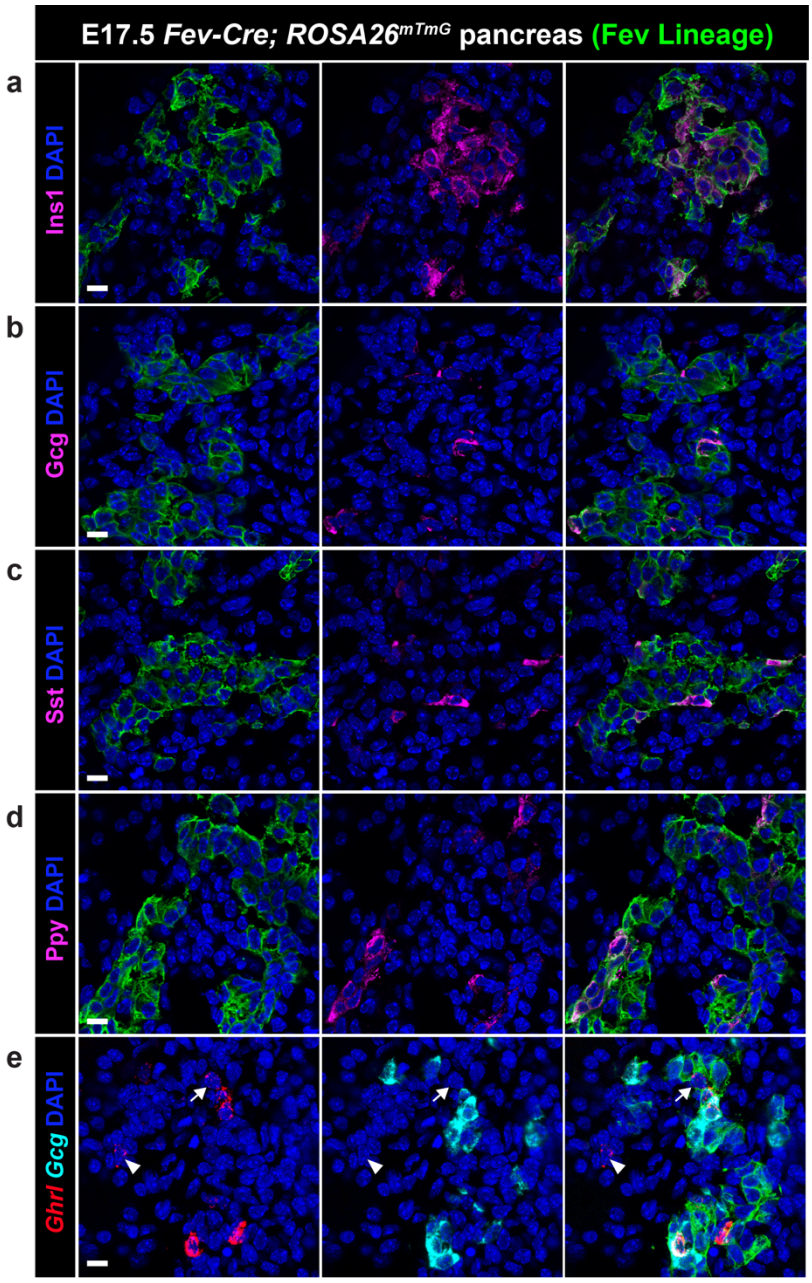


Figure 2.12. *In vivo* Fev lineage tracing of E17.5 mouse pancreata. (a-d) Immunofluorescence (IF) for hormones Ins1 (100% lineage-labeled), Gcg (100% lineage-labeled), Sst (96.7% lineage-labeled), and Ppy (100% lineage-labeled) in embryonic pancreatic hormones in *Fev-Cre; ROSA26*^{mTmG} lineage traced animals at E17.5 (n=86 cells of 5 pancreata for Ins1; n=57 cells of 5 pancreata for Gcg; n=30 cells of 5 pancreata for Sst; n=47 cells of 5 pancreata for Ppy). (e) Dual IF (for membrane-GFP) and multiplexed fluorescent ISH for *Ghrl* and *Gcg* in *Fev-Cre; ROSA26*^{mTmG} lineage traced animals at E17.5 (n=23 cells of 2 pancreata for *Ghrl/Gcg*). *Ghrl*+/*Gcg*- cells (47.8% lineage-labeled) represent the epsilon population. Non-lineage labeled epsilon cells are denoted by the arrowheads, and lineage-labeled epsilon cells are denoted by the arrows. Scale bar represents 10um in a-e.

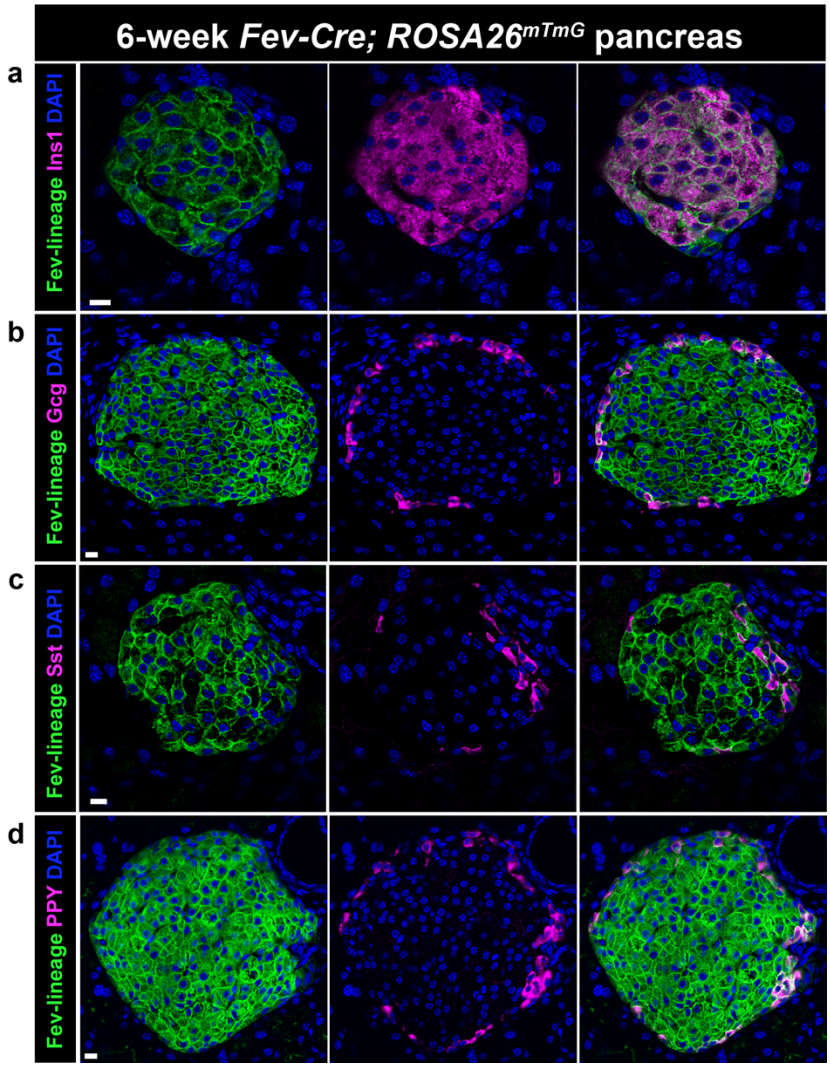


Figure 2.13. *In vivo* Fev lineage tracing of adult mouse pancreata. (a-d) IF for adult hormones in 6-week *Fev-Cre; ROSA26*^{mTmG} lineage-traced pancreas. From one animal: n=172 cells for Ins1 (100% lineage-labeled); n=65 cells for Gcg (100% lineage-labeled); n=86 cells for Sst (97.7% lineage-labeled); n=30 cells for Ppy (100% lineage-labeled). Scale bar represents 10um in a-d.

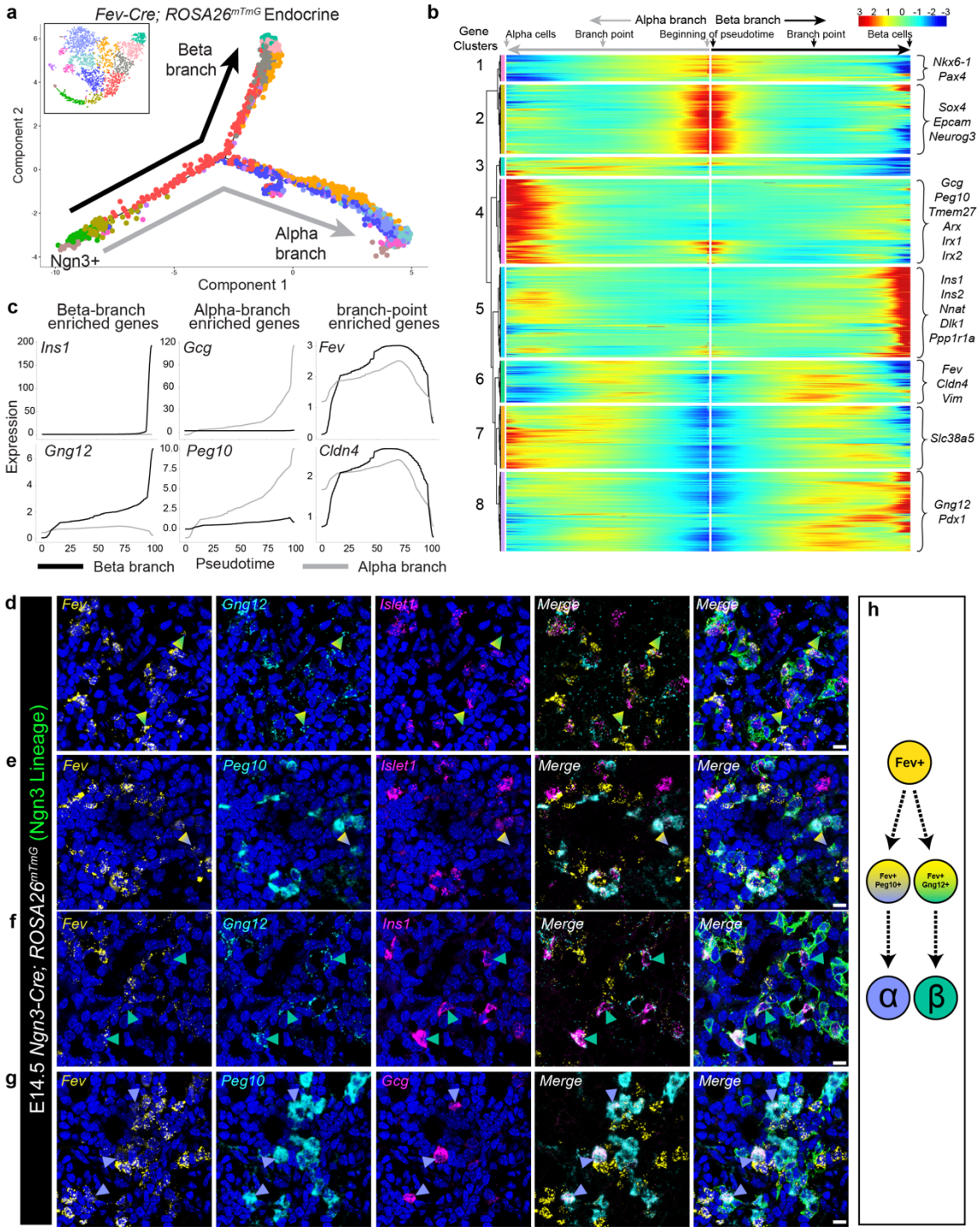


Figure 2.14. Identification of candidate regulators of beta and alpha cell fate decisions. (a) Pseudotime ordering of the endocrine cells at E14.5 depicted in Fig. 2.8h yields a bifurcated tree in which the two main branches terminate in cells that highly express *Ins1* (beta cell branch) or *Gcg* (alpha cell branch). (b) Heatmap depicting the expression of genes along each branch, in pseudotime. An independent expression pattern is calculated across the entire pseudotime trajectory for each branch. Therefore, the portion of the trajectory before the branch point is displayed for each branch separately. Genes are clustered based on expression pattern across

pseudotime; selected genes with differential expression along the branches are highlighted to the right. (c) Gene expression plots depicting the kinetic trends along each branch. (d-e) Multiplexed fluorescent ISH for *Fev*, *Gng12*, and *Islet1* (d) or *Fev, Peg10,* and *Islet1* (e) in lineage-traced E14.5 *Ngn3-Cre; ROSA26*^{mTmG} pancreas. Arrowheads identify lineage-traced Fev+/Islet1- cells with *Gng12* (d, teal gradient arrowheads) or *Peg10* (e, indigo gradient arrowheads) expression. (f) Multiplexed fluorescent ISH for *Fev*, *Gng12*, and *Ins1*. Teal arrowheads identify lineage-traced Ins1+ beta cells that express *Gng12*. (g) Multiplexed fluorescent ISH for *Fev*, *Peg10*, and *Gcg*. Indigo arrowheads identify lineage-traced Gcg+ alpha cells that express *Peg10*. (h) Model for Fev^{Hi} (yellow) cell differentiation into distinct alpha or beta cells. *Peg10* and *Gng12* expression in Fev^{Hi} cells may represent progenitors pre-fated towards the alpha and beta lineages, respectively, during endocrine lineage allocation. (d-g) Scale bars represent 10 um. Blue staining represents DAPI-labeled nuclei. Colors of arrowheads match colors of cells represented in (h).

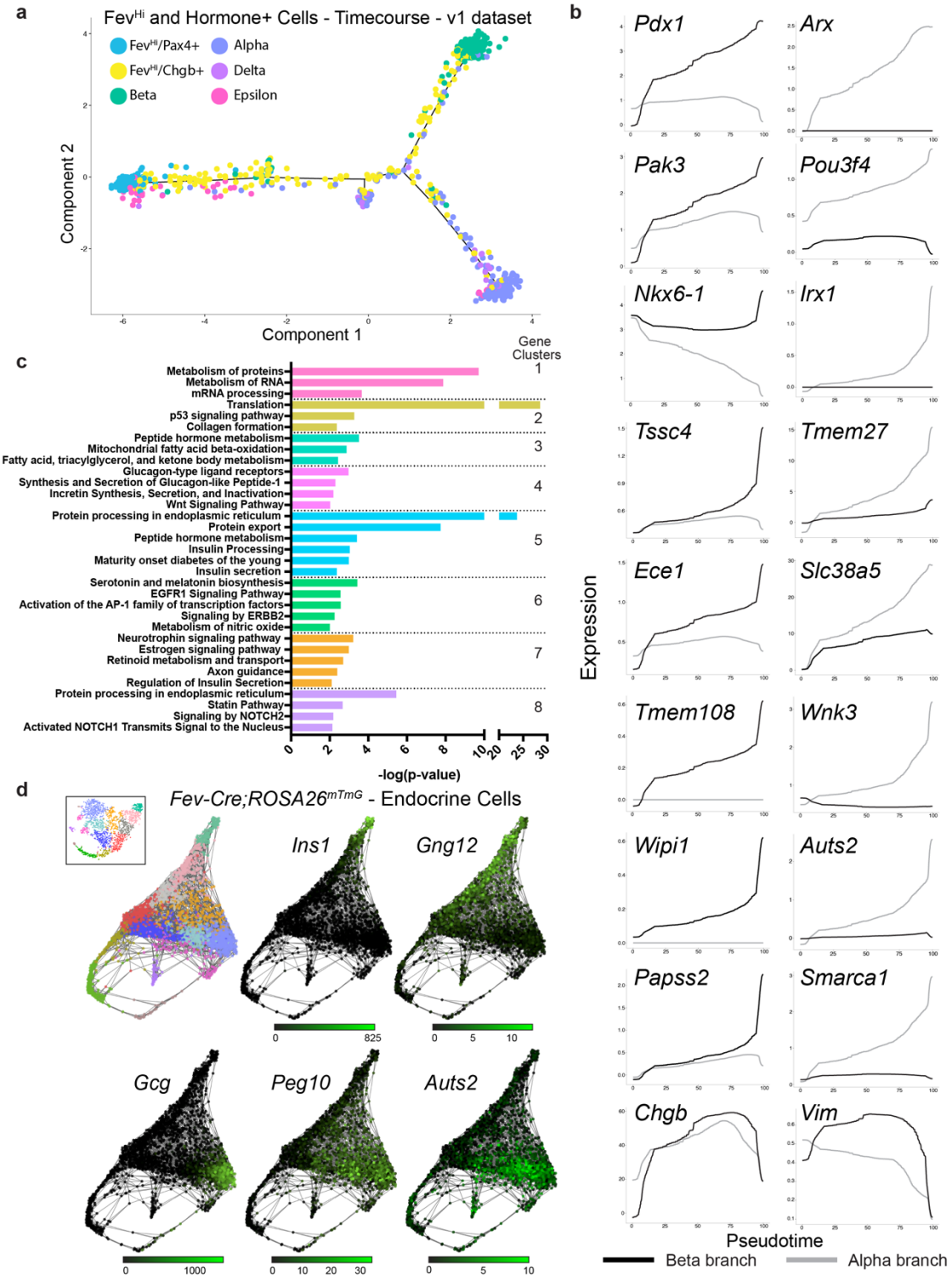


Figure 2.15. Identification of Candidate Genes and Pathways Enriched Along Beta and Alpha Cell Lineages. (a) Pseudotime ordering trajectory of v1 timecourse dataset, including E12.5, E14.5 (batch 1 and batch 2), and E17.5 datasets. (b) Gene expression plots depicting the kinetic curves of individual genes (from Fig. 2.14b) across pseudotime in the alpha or beta branches. (c) Pathway analysis for clusters of genes from the BEAM analysis. Gene clusters

correspond to Figure 14b. (d) SPRING plots for Fev-lineage traced dataset, including all endocrine cells. Colors match those in Fig. 2.11h. Expression of selected genes predicted from monocle BEAM analysis.

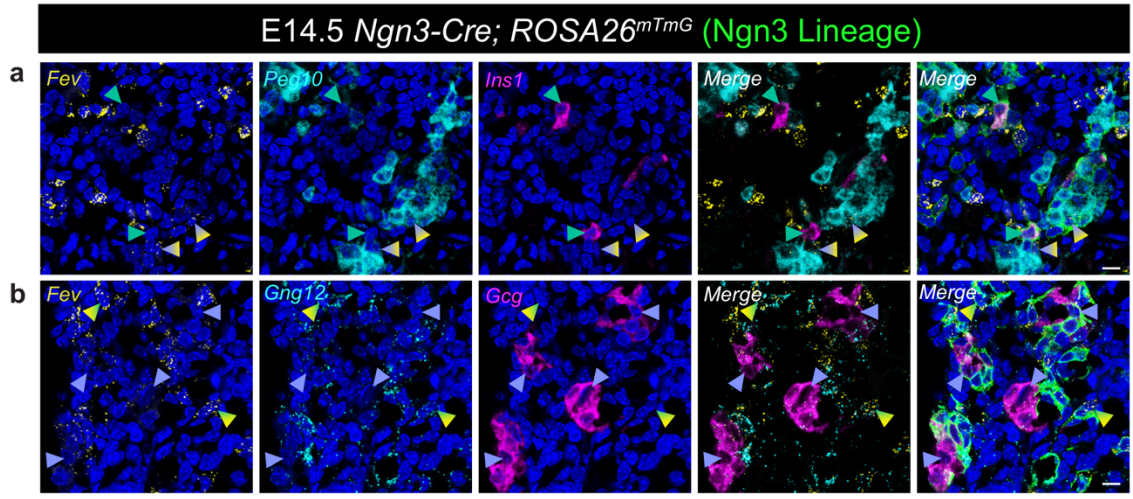


Figure 2.16. Expression of Candidate Regulators within the Endocrine Lineage Prior to Alpha or Beta Cell Identity. (a) Multiplex fluorescent ISH for *Fev* (yellow), *Peg10* (cyan), and *Ins1* (magenta) in lineage-traced E14.5 *Ngn3-Cre; ROSA26*^{mTmG} pancreas. Indigo-graded arrows highlight lineage-traced *Fev*+/*Peg*10+ cells that do not express *Ins1*. Teal arrows highlight *Ins1*+ beta cells that do not express *Peg10*. (b) Multiplex fluorescent ISH for *Fev* (yellow), *Gng12* (cyan), and *Gcg* (magenta) in lineage-traced E14.5 *Ngn3-Cre; ROSA26*^{mTmG} pancreas. Teal-graded arrows highlight lineage-traced *Fev*+/*Gng12*+ cells that do not express *Gcg*. Indigo arrows highlight *Gcg*+ alpha cells that are not enriched for *Gng12*.

Chapter 3

Lineage allocation in human endocrine cell development

Introduction

Defining endocrine lineage allocation in mouse pancreatic development has been largely motivated by the desire to understand how endocrine cells arise in human development. Gaining a better understanding of the cell stages required for human endocrine cell development as well as the transcriptional circuitry driving lineage allocation into distinct endocrine linages will refine our ability to generate these hormone-expressing cell types from human embryonic stem cells. Several key transcription factors important for endocrine cell development in mouse have already been found to be conserved and required in that of human. Like in *Pdx1*-null mice, inactivating mutations in *PDX1* results in neonatal diabetes that requires insulin treatment as well as exocrine pancreas deficiency (Stoffers et al., 1997). *PDX1* is expressed in the pancreatic endoderm and is retained in multipotent pancreatic progenitors that are postulated to give rise to acinar, duct, and endocrine lineages of the human pancreas (Jennings et al., 2013; Lyttle et al., 2008; Piper et al., 2004). Homozygous mutations in *NGN3* also cause neonatal diabetes due to the lack of differentiated beta cells in the pancreas. This pancreatic phenotype of *NGN3*-deficient neonates is also accompanied with the lack of intestinal enteroendocrine differentiation that results in congenital diarrhea (Pinney et al., 2011; Wang et al., 2006), which is concordant with Ngn3's role in enteroendocrine differentiation in murine intestine (Gehart et al., 2019). Downstream of *NGN3*, the transcription factors *NEURDO1*, *NKX2.2*, *PAX6*, and *MNX1* have all been identified to cause neonatal diabetes when mutated (Bonnefond et al., 2013; Flanagan et al., 2014; Rubio-Cabezas et al., 2010; Solomon et al., 2009). *NEUROD1* and *NKX2.2* are considered direct targets of *NGN3* (Gradwohl et al., 2000; Sussel et al., 1998) and, like PAX6, are expressed in differentiated beta cells in early development (Jennings et al., 2013; Lyttle et al., 2008). Thus, there is functional conservation of critical transcription factors required for pancreatic development in both mouse and human.

In particular, although the function of *NGN3* as a key regulator in endocrine cell development appears to be conserved in human, the duration of *NGN3* expression and the events following *NGN3* upregulation in human pancreatic development remain unclear. In mouse, *Ngn3* expression in the pancreas is biphasic (Villasenor et al., 2008) and within each endocrine progenitor, *Ngn3* expression persists for about 24 hours (Beucher et al., 2012b). In human, it has been reported that NGN3 expression occurs as a single wave (Jennings et al., 2015). However, the duration of this single wave of *NGN3* expression in human development has not been fully mapped, and the lack of genetic and time-resolution studies of human endocrine development makes it difficult to assess the true dynamics of *NGN3* expression. Additionally, in human, hormone-expressing lineages are presumed to be derived from NGN3+ endocrine progenitors, but the events following transient *NGN3* expression and the differentiation into a hormone-expressing endocrine lineage remain unclear. Without understanding the cell stages during this critical window of endocrine cell differentiation, we also lack understanding of the transcriptional mechanisms that may be regulating endocrine lineage allocation in one of five distinct hormone-expressing lineages in human pancreatic development.

The discovery of a novel endocrine progenitor defined by high *Fev* expression in mouse prompted the question of whether additional endocrine progenitor stages exist in human endocrine cell development beyond the NGN3+ endocrine progenitor stage. Defining these stages in human can be leveraged to more properly mimic human endocrine cell development *in vitro* during directed differentiation protocols that harness the power of hESCs. Here, we use single-cell RNA-sequencing to interrogate the diversity of cell types that arise during human pancreatic development. We construct transcriptional profiles of previously known human pancreatic cell compartments, including epithelial, mesenchymal, endothelial, neural, and immune populations. Within the endocrine compartment, we identify previously known NGN3+ endocrine progenitors and uncover new transcriptional markers of these cells. We also annotate alpha, beta, delta, and epsilon lineages based on expression of *GCG, INS, SST*, and *GHRL*,

respectively. *In silico* lineage reconstruction of endocrine lineages identifies two novel endocrine populations in human pancreatic development that represent pre-beta and pre-alpha endocrine progenitors derived from NGN3+ progenitors and that distinctly precede differentiation into the alpha and beta lineages. Both these pre-beta and pre-alpha progenitor populations express *FEV*. This *in silico* reconstruction of endocrine lineage relationships also enabled the identification of candidate regulators of alpha and beta cell lineage fates in human development. Using *in vivo* human beta cell development as a guidepost for proper beta cell generation from *in vitro* differentiation of hESCs, we construct a lineage map of hESC-derived beta cell differentiation and identify *FEV*-expressing endocrine cells that have mis-differentiated into a hormone-negative lineage. Dissection of the transcriptomic profiles of this mis-differentiated lineage reveals candidate regulators of improper beta cell differentiation that become expressed during the directed differentiation process. Thus, we now have mapped proper and improper differentiation trajectories of human endocrine lineages during both fetal development and the directed differentiation of hESCs towards a beta lineage. Our work forms a roadmap from which our understanding of human endocrine cell development can be further refined.

Results

Diversity of cell types in the developing human pancreas

We first characterized the different cellular compartments and their transcriptional profiles from human fetal pancreas. We focused on a 12wpc timepoint, which represents a period of peak NGN3 expression and active cell differentiation in the developing human pancreas (Nair and Hebrok, 2015). Tissue from this 12wpc timepoint was dissociated into a single-cell suspension, and RBCs were removed via immunomagnetic separation. The resulting single-cell suspension was loaded onto two wells of the 10X Chromium Single-Cell Platform and prepared for sequencing using version 3 (V3) chemistry. Following sequencing and de-

multiplexing of single-cell data, UMAP-based clustering of merged well replicates revealed 22 cell clusters organized into acinar, ductal, endocrine, mesenchymal, endothelial, immune, and nerve populations based on the expression of known marker genes, such as *CPA1* (acinar), *SOX9* (ductal), *CHGA* (endocrine), *COL1A1* (mesenchymal), *PECAM1* (endothelial), *PTPRC* (immune), and *SOX10* (nerves) (Fig. 3.1a-c). The contribution of each cell cluster from each replicate well showed no batch effect based on use of different lanes on the Chromium 10X platform (Fig. 3.1a). Similar to our work using single-cell RNA-sequencing to interrogate the diversity of cell populations in mouse pancreatic development, transcriptomic profiling at single-cell resolution facilitated our preliminary construction of a cellular roadmap of the cell populations present in the developing human pancreas.

Identification of novel cell stages during human endocrine cell development

Given our previous identification of novel progenitor stages in mouse pancreatic development, we next focused on the endocrine compartment of the developing human pancreas in order to determine if additional endocrine stages exist beyond those characterized by NGN3+ progenitors and differentiated hormone+ endocrine cells. Sub-clustering of the CHGA+ cell clusters resulted in increased resolution of the endocrine lineage populations present in 12wpc human fetal pancreas, revealing 11 distinct endocrine lineage populations (Fig. 3.2a). We identified clusters defined by *GCG*, *INS*, *SST*, and *GHRL*, corresponding to alpha, beta, delta, and epsilon cells, respectively (Fig. 3.2b). By sequencing, *INS* transcript appeared to be also present in the alpha cell cluster (Fig. 3.2b). However, it is likely that differentiated alpha cells do not translate the *INS* transcript into protein. While we only identified one cell cluster each for alpha, delta, and epsilon populations, we observed 3 distinct fetal INS+ beta cell populations (clusters 0, 2, and 4) defined by distinct differentially-expressed genes (Fig. 3.2a-c). Many of these differentially-expressed genes, such as *DLK1*, *MEG3*, and *RBP4*,

have been previously identified to drive beta cell heterogeneity in the adult human pancreas, suggesting that sources of beta cell heterogeneity arise as early as in fetal pancreatic development (Lawlor et al., 2017; Segerstolpe et al., 2016). This heterogeneity in beta cells observed in development could serve as a major underpinning of the heterogeneity in the regulation of functional maturation and levels of ER stress seen in both mouse and human beta cells (Baron et al., 2016; Muraro et al., 2016; Qiu et al., 2017a; Zeng et al., 2017).

We also found three additional cell clusters (clusters 6, 8, and 9) that were devoid of any hormone expression (Fig. 3.2a,b and 3a). Cluster 6 displayed expression of *NGN3* (Fig. 3.2b, 3a), and clusters 8 and 9 exhibited *FEV* expression (Fig. 3.2b, 3a). Given our previous work identifying that *Fev* marks a endocrine progenitor population that is derived from Ngn3+ progenitors in mouse pancreatic development (Byrnes et al., 2018), we wanted to determine whether these clusters gave rise to hormone-expressing endocrine lineages. We performed *in silico* reconstruction of lineage relationships using pseudotemporal ordering with only cluster 6, cluster 8, cluster 9, beta 1, beta 2, and the alpha cell cluster as input (Fig. 3.3b, left panel). This analysis revealed that clusters 6, 8, and 9 were precursors to the alpha and beta lineages (Fig. 3.3b, c). Cluster 6 represented a common endocrine progenitor population that gave rise to both alpha and beta lineages (Fig. 3.3a, b). Clusters 8 and 9, on the other hand, were precursor populations that gave rise to differentiated beta and alpha cells, respectively (Fig. 3.3a-c). Thus, cluster 8 was postulated to represent a pre-beta progenitor population, and cluster 9 was postulated to represent a pre-alpha progenitor population. *FEV* was a top 1.5-fold differentially expressed gene in both clusters 8 and 9 (Fig. 3.3a and Supplementary Data 17), suggesting that FEV-expressing progenitors give rise to both alpha and beta lineages in human endocrine cell development. *FEV* expression persisted into the alpha lineage but was not expressed by differentiated beta cells (Fig. 3.3a, c), suggesting that *FEV* must turn off in pre-beta progenitors prior to the acquisition of beta cell identity while this requirement is not true for pre-alpha progenitors that differentiate into the alpha lineage. This was in contrast to mouse pancreatic

development in which *Fev* is expressed in a subset of differentiated alpha and beta cells (Byrnes et al., 2018). In examining other critical transcription factors for endocrine development, we observed upregulation of *PDX1*, *NKX6.1*, and *PAX4* expression in the pre-beta progenitor stage (Fig. 3.3c). *PAX6* and *NEUROD1*, on the other hand, were expressed in both the alpha and beta lineages, starting at the common progenitor stage (cluster 6) (Fig. 3.3c). Our pseudotemporal ordering analysis thus provides additional insight into the cell stages required to generate alpha and beta cells in human pancreatic development.

The inclusion of other hormone-expressing endocrine lineages in pseudotemporal ordering did not result in a continuous differentiation trajectory (Fig. 3.3b, middle and right panels). Inclusion of the third beta cell population (beta 3) resulted in a trajectory that excluded this population from the main trajectory (Fig. 3.3b, middle panel), meaning that our pseudotemporal ordering analysis considered this beta 3 cluster as not lineage related to any other population. Similarly, inclusion of the GHRL+ epsilon and SST+ delta clusters resulted in a disjointed trajectory that did not place these other hormone-expressing lineages in lineage relationships with the progenitor populations we have annotated (Fig. 3.3b, right panel). Sequencing of additional timepoints will increase the number of endocrine lineage cells captured and thus may provide increased resolution to differentiation process of these other hormone-expressing lineages.

Pairwise comparisons and examination of the top differentially expressed genes of clusters 6, 8, and 9 revealed that these populations represent novel cell stages of human endocrine cell differentiation at a resolution that we have not been able to appreciate with previous techniques. *NGN3* expression was concentrated within the common endocrine progenitor population (cluster 6), although *NGN3* was not among the top 5 differentially-expressed genes (Fig. 3.2b and Supplementary Data 17). Instead, this common endocrine progenitor cluster was defined by expression of genes such as *EMC10, SOX4*, and *HES6* (Fig. 3.2c, 4a-c). EMC10 is a ER membrane protein with reported roles as an angiogenic factor that

promotes tissue repair after myocardial infarction (Reboll et al., 2017). SOX4 is a member of the SOX family of transcription factors and is a reported target of *Ngn3* in mouse pancreatic development (Xu et al., 2015). HES6 suppresses HES1, which suppresses the onset of *NGN3* expression that initiates endocrine cell development in the pancreas (Masjkur et al., 2016). Other notable genes whose canonical functions involve transcription and DNA binding and that were more than 1.5-fold differentially expressed in common endocrine progenitors (cluster 6) include *NEUROD1, NKX2-2, PAX4, RFX3, SMARCC1, CITED2, HMGB3, KDM5B, ZBTB18, TGIF2, ARID4A, CBFA2T2,* and *PROX1* (Fig. 3.4 and Supplementary Data 17). According to our pseudotime reconstruction of endocrine lineage relationships, these common endocrine progenitors in cluster 6 gave rise to both pre-beta and pre-alpha progenitors (clusters 8 and 9) (Fig. 3.3b). Although both the pre-beta and pre-alpha progenitor populations expressed hormones (Fig. 3.3a), their expression of either *INS* or *GCG* was markedly lower compared to the expression levels of both hormones found in differentiated beta or alpha cells (Fig. 3.3a), suggesting that these clusters were not fully differentiated into their respective endocrine lineages. The top differentially-expressed genes in pre-beta progenitors in cluster 8 were *MEG3*, *NR4A2*, and *IGFBP5* (Fig. 3.2c). Notable factors involved in transcription that were more than 1.5-fold differentially-expressed in pre-beta progenitors include *SOX4, NKX6-1, PDX1, PAX6, PAX4, EGR3, ARID5B, RYBP, SIM1, MNX1, TSHZ1, ATF3, FOXA2, NR4A3, NR4A1, NR4A2, MAFB, EGR4, NPAS4, ID4, and ETS2* (Fig. 3.4a, b, and d and Supplementary Data 17). In contrast, the top differentially expressed genes in pre-alpha progenitors in cluster 9 were *IRX2* and *ARX*, which both regulate alpha cell lineage allocation (Petri et al., 2006; Wilcox et al., 2013), as well as *CDKN1C*, which is a cyclin-dependent kinase inhibitor (Fig. 3.2c). Notable factors involved in transcription that were more than 1.5-fold differentially-expressed in pre-alpha progenitors include *NEUROD1, PAX6, ISL1, PSIP1, ST18, SIM1, MLXIPL, TOX3, PBX1, ESRRG, and ID4* (Fig. 3.4a, c and d and Supplementary Data 17). Pairwise comparisons

among these three endocrine progenitor clusters reveal that each population is transcriptionally distinct and arises at defined stages along endocrine cell differentiation.

Candidate lineage regulators of the beta cell lineage

The onset of *NGN3* expression marks the beginning of endocrine cell development as cells differentiate towards a hormone+ endocrine lineage. However, the transcriptional programs that guide these endocrine progenitors toward a distinct hormone-expressing endocrine lineage are not well defined in human endocrine cell differentiation. With the lack of lineage tracing tools available for *in vivo* human studies, we used our single-cell RNA-sequencing data to make inferences about the transcriptional machinery that regulates endocrine lineage allocation. Given that we observed distinct stages of cellular differentiation leading to both alpha and beta lineages (Fig. 3.3b), we next investigated potential transcriptional regulators involved in mediating the acquisition of either alpha or beta cell identity.

We used pseudotemporal ordering to identify genes that were differentially expressed across a single-cell trajectory. We first applied this analysis to the beta lineage branch, which exhibited differentiation starting with cluster 6 cells (common endocrine progenitors) to cluster 8 cells (pre-beta progenitors) and finally to beta cells (Fig. 3.5a). Our lineage branch analysis resulted in 7 major gene clusters that displayed 3 main patterns of gene expression: genes that were highly expressed in the common progenitor stage but tapered in expression as differentiation proceeds (gene clusters 2-4), genes that turned on specifically in the pre-beta progenitor stage and were either subsequently downregulated or remained expressed (gene clusters 1, 6, and 7), and genes that turned on specifically in differentiated beta cells (gene cluster 5) (Fig. 3.5a). *NGN3* was found in gene cluster 3, along with other genes known as endocrine progenitor markers, such as *NKX2-2, RFX6, NEUROD1, PROX1, HES6, and GATA6* (Fig. 3.5a and Supplementary Data 18). These known endocrine progenitor markers were highly

expressed in the common progenitor cluster (cluster 6) but were downregulated during the pre-beta progenitor stage (Fig. 3.5b). On the other hand, genes within gene cluster 7, which were expressed specifically in differentiated beta cells, included *WNT4*, which regulates beta cell proliferation, and *RGS2*, which regulates beta cell mass (Supplementary Data 18) (Dong et al., 2017; Heller et al., 2011). *INS*, which is a definitive marker of beta cells, began to become upregulated in the pre-beta progenitor stage and reached peak expression in the differentiated beta cell stage (Fig. 3.5c).

We reasoned that genes within gene clusters 1, 6, and 7 that were upregulated during the pre-beta progenitor stage may serve as potential regulators of beta cell lineage allocation. *FEV* was found in gene cluster 7 (Supplementary Data 18). Genes that displayed high upregulation during the pre-beta progenitor stage also included genes known to be involved in beta cell differentiation and function, such as *CHGB, SCG5, ERO1B, MAFB,* and *PAX6* (Fig. 3.5c). The upregulation of known genes involved in insulin production, insulin processing, and the beta cell differentiation program confirmed that our pseudotemporal ordering analysis was robust. Genes involved in cytoskeletal remodeling and cell migration, including *ROBO2, VIL1, CALY, TAGLN2, KIF5C, KIF12,* and *PHACTR2,* were also upregulated beginning in the pre-beta progenitor stage (Supplementary Data 18). These cytoskeletal remodeling and cell migration genes could be reflective of islet cell migration and formation that occurs concurrently with beta cell differentiation (Sharon et al., 2019a).

We also identified candidate regulators not previously known to be involved in beta cell lineage allocation. These genes were organized into 3 broad categories: those that were imprinted, those involved in neural development, and others involved in transcription and canonical signaling pathways. Imprinted genes that were upregulated during beta cell differentiation included *DLK1, MEG3, GNAS, PLAGL1, PEG3,* and *PEG10* (Fig. 3.5d). The role of imprinting in endocrine cell development has not been previously explored, although loss or dysregulation of imprinted genes has been implicated in impaired pancreatic endocrine cell

function. Epigenetic dysregulation of *DLK1-MEG3* microRNAs has been observed in human T2D islets (Kameswaran et al., 2014). In the *GNAS* locus, improper mono-allelic expression of *GNAS* from the maternal allele results in a host of growth and metabolic disorders, including obesity (Weinstein et al., 2010). Bi-allelic expression of *PLAGL1* causes 60% of all cases of transient neonatal diabetes mellitus (TNDM) (Hoffmann and Spengler, 2012; Kamiya et al., 2000). *PEG3* encodes a zinc finger transcription factor that inhibits beta cell proliferation in mice (Sojoodi et al., 2016), and *PEG10* was recently identified by our group to be a lineage marker of the murine alpha lineage (Byrnes et al., 2018).

Upregulated genes known to be involved in neural development included *ASCL2*, *AHI1*, and *SEZ6L2* (Fig. 3.5e). *ASCL2* (also called *MASH2*) is also an imprinted gene and comes from a family of bHLH transcription factors that regulates neuronal progenitor differentiation and peripheral nerve regeneration (Ge et al., 2006; Guillemot et al., 1993; Küry et al., 2002). *AHI1* regulates cortical development in humans (Doering et al., 2008). Mutations in *SEZ6L2* have been implicated in seizure-related phenotypes, and more recently, the gene has been identified as marker for developing islet cells during embryogenesis (Bedoyan et al., 2010; Hald et al., 2011). Finally, *ARID5B* and *ACVR1C* were genes significantly expressed as differentiation into the beta lineage occurred (Fig. 3.5e). *ARID5B* participates as a transcriptional coactivator that is required for adipogenesis (Okuno et al., 2013). Signaling through *ACVR1C*, an Activin A receptor, has been reported to inhibit insulin secretion from beta cells (Bertolino et al., 2008), suggesting that signaling downstream of *ACVR1C* possibly blocks premature insulin secretion in pre-beta progenitors or early-forming beta cells during human endocrine cell development. The identification of these genes upregulated during the pre-beta progenitor stage begins to elucidate transcriptional circuitry involved in human beta cell differentiation that can be subsequently validated in *in vitro* differentiation models (Fig. 3.5f).

Candidate lineage regulators of the alpha cell lineage

We applied a similar analysis to identify genes that were differentially expressed during differentiation into alpha cells. Based on pseudotemporal ordering, differentiation of the alpha lineage from endocrine progenitors began with cluster 6 (common progenitors) that differentiate into cluster 9 cells (pre-alpha progenitors), which then become differentiated alpha cells found in cluster 9 (Fig. 3.3b). Alpha lineage branch analysis resulted in 6 major gene clusters that, similar to the beta lineage branch analysis (Fig. 3.5a), displayed 3 main patterns of gene expression: genes that were highly expressed in the common progenitor stage but tapered in expression as differentiation proceeds (gene clusters 2 and 3), genes that turned on specifically in the pre-alpha progenitor stage and were either subsequently downregulated or remained expressed (gene clusters 1, 4, and 5), and genes that turned on specifically in differentiated alpha cells (gene cluster 6) (Fig. 3.6a). Genes known to be expressed by differentiated alpha cells, including *GCG, TTR, ALDH1A1, FAM46A*, and *CRYBA2* (Dorajoo et al., 2017; Muraro et al., 2016; Su et al., 2012), displayed upregulated expression along pseudotime (Fig. 3.6b). Genes known to regulate alpha lineage allocation, such as *IRX2, ARX, and ISL1*, also were upregulated but specifically beginning in the pre-alpha progenitor cluster (Fig. 3.6c).

Given that gene clusters 4 and 5 contained genes that were upregulated specifically after NGN3 downregulation and the acquisition of alpha cell identity, these genes may potentially serve as candidate regulators of alpha lineage allocation. Similar to our analysis of the beta cell lineage, *MAFB, PAX6, ERO1B, AHI1, PEG10, SCG5,* and *ACVR1C* were found to be upregulated along alpha cell differentiation (Supplementary Data 19), suggesting that these markers are common genes upregulated during endocrine cell differentiation as a whole. We observed upregulation of various neural transcription factors and genes during alpha cell differentiation. *BEX2, BEX4,* and *BEX5* were all upregulated during alpha cell fate allocation and are members of the brain-expressed X-linked transcription factor family that are highly expressed in the brain (Alvarez et al., 2005) (Fig. 3.6d). *ST18* was another neuronal lineage

transcription factor that promotes cholinergic motor neuron differentiation and that was upregulated during alpha cell development (Teratani-Ota et al., 2016) (Fig. 3.6d). Other genes upregulated during this process and implicated in neural function included *ANK3*, which is required in neurons for proper synapse structure and function (Smith et al., 2014), and *STMN2*, which is required for normal axonal outgrowth and regeneration in the nervous system (Klim et al., 2019) (Fig. 3.6d). Furthermore, we identified a number of cell surface markers that were first expressed during the pre-alpha progenitor stage and can be used to guide cell purification. These included *SLC3A2, SLC7A2, SLC7A8, SLC30A8, ALCAM,* and *CD99* (Fig. 3.6e). SLC30A8 has already been shown to be required in adult alpha cells for hypoglycemia-induced glucagon secretion (Solomou et al., 2015). Our analysis of the alpha differentiation trajectory (Fig. 3.6f), paired with that of the human beta lineage, highlights the power of pseudotemporal ordering that defines the dynamic transcriptional programs in place as endocrine progenitors become specified towards distinct hormone-expressing lineages.

Cellular and transcriptional dynamics of the developing endocrine compartment

To understand endocrine cell development across actual developmental time, we performed single-cell RNA-sequencing on tissue at 3 additional timepoints to add to our analysis on 12wpc human fetal pancreas (referred to as 12wpc_1): a second biological replicate of 12wpc (referred to as 12wpc_2), 15.5wpc, and 16wpc. All datasets were depleted of red blood cells through immunomagnetic separation and generated using the 10X Genomics version 3 (V3) sequencing chemistry, except for the 15.5wpc sample, which was enriched for EPCAM+ cells through FACS and processed using version 2 (V2) sequencing chemistry. All datasets were merged with Seurat 3's new integration method for merging and batch correction, resulting in 31 distinct clusters (Fig. 3.7a) that were classified as CPA1+ acinar cells, SOX9+ ductal cells, CHGA+ endocrine cells, COL1A1+ mesenchymal cells, PECAM1+ endothelial cells, PTPCR+ immune cells, and SOX10+ nerve cells (Fig. 3.7c). The two 12wpc samples and single 16wpc

sample were represented in all clusters, highlighting the robustness of Seurat 3's new Integration method for batch correction by biological sample (Fig. 3.7b). As expected, given that the 15.5wpc sample was enriched for EPCAM+ cells, this timepoint only contributed to the EPCAM+ populations in the merged dataset (Fig. 3.7b, c). To focus on the endocrine compartment, CHGA+ endocrine clusters from the merged dataset were sub-clustered (Fig. 3.7c).

Sub-clustering of the endocrine lineage resulted in 15 distinct populations (Fig. 3.8a), including NGN3+ endocrine progenitors, INS+ beta cells, GCG+ alpha cells, SST+ delta cells, and GHRL+ epsilon cells (Fig. 3.8c). We also observed a FEV+ cluster that was not defined by any hormone expression (Fig. 3.8c), which is in line with the FEV+ endocrine progenitor cells we identified in 12wpc pancreas (Fig. 3.2b, 3a). This FEV+ cluster had representation from all 4 fetal timepoints (Fig. 3.8b, c), suggesting that FEV+ progenitors appear as early as 12wpc and persist at least as late as 16wpc in human pancreatic development. Plotting the top 3 differentially-expressed genes from each cluster highlighted the gene expression profile differences of each endocrine cluster (Fig. 3.8d).

We next sought to reconstruct lineage relationships across multiple timepoints through pseudotemporal ordering. Batch effect, unfortunately, is a major issue that still confounds single-cell RNA-sequencing analysis, despite multiple groups developing algorithms to address this problem (Butler et al., 2018; Haghverdi et al., 2018; Stuart et al., 2019). In our merged dataset, batch effect became problematic as our 15.5wpc sample was processed with V2 10X Genomics version chemistry as opposed to V3 chemistry, which was utilized for processing our 12wpc and 16wpc samples. V3 chemistry increases the sensitivity of gene capture, and this was particularly evident by the percentage of mitochondrial genes captured in V3 datasets, in which more mitochondrial genes were represented in V3 datasets (Fig. 3.8e). This discrepancy in mitochondrial gene capture led to a batch effect in pseudotemporal ordering that was not resolved through regressing on mitochondrial content (Fig. 3.8e). Endocrine cells from the

15.5wpc sample clustered more closely to one another than integrating with cells from other timepoints, and this was driven by differences in mitochondrial content of the datasets (Fig. 3.8e). In order to reconstruct lineage relationships across multiple timepoints, additional human fetal datasets using V3 chemistry will be used in the future.

Understanding the emergence of distinct cellular compartments during *in vitro* beta cell differentiation at single-cell resolution

Directed differentiation of hESCs to a beta cell lineage represents a powerful approach for not only generating beta cells for diabetes but also understanding human beta cell differentiation. Given the significant heterogeneity in cells generated by directed differentiation of hESCs towards the beta lineage, we leveraged single-cell RNA-sequencing to classify distinct cellular populations that arose across five main stages of *in vitro* beta cell differentiation: stages containing early-, middle-, and late-stage endocrine progenitors (ES4, S5D4, and S5D7) and two stages within the beta lineage stage (S6D4 and S6D10). UMAP-based clustering of all five timepoints revealed the presence of PDX1+ clusters, reflecting induction towards the pancreatic lineage during the directed differentiation towards the beta lineage (Figs. 9, 10). In ES4, we identified proliferating PDX1+ pancreatic progenitors, PDX1+/NKX6.1+ pancreatic progenitors, early-induced endocrine cells marked by *CHGA* and *NEUROD1*, and CDX2+ clusters that likely represent intestinal lineages that arose from improper differentiation (Fig. 3.9c). Of mid- to late-stage endocrine progenitors in Stage 5, only a small percentage expressed the endocrine progenitor marker *NGN3* (Fig. 3.9d, e), which was expected given the transient nature of *NGN3* expression. Instead, we observed PDX1+/NKX6.1+ progenitors, replicating PDX1+/NKX6.1+ progenitors, the persistence of CDX2+ cluster, and CHGA+/NEUROD1+ endocrine clusters that began to express hormones, such as *INS* and *GCG* (Fig. 3.9d, e). During Stage 6, which is defined as the beta cell stage, we again observed CDX2+ clusters and CHGA+/NEUROD1+ endocrine clusters that contain INS- and GCG-producing cells (Fig. 3.10a-d).

hESC-derived FEV+ cells are transcriptionally similar to *in vivo* FEV+ progenitors

Given that we had identified an endocrine progenitor stage defined by *FEV* expression in both mouse (Byrnes et al., 2018) and human fetal beta cell development (Fig. 3.2b), we examined whether the derivation of hESC-derived beta cells also involved transit through a *FEV*-expressing cell stage. From qPCR, we determined that *FEV* began to be expressed starting in Stage 4 pancreatic progenitors and was robustly expressed in Stage 5 endocrine progenitors (Fig. 3.11a). *FEV* expression persisted in cells at the Stage 6 beta cell stage (Fig. 3.11a), which was in contrast to differentiated beta cells in human pancreatic development that did not express *FEV* (Fig. 3.2b, 3a). This was consistent with *in situ* hybridization for *FEV*, in which *FEV* transcript was observed in S5D3 and S6D11 clusters (Fig. 3.11b). Similarly, in our transcriptomic profiling of cells from the end of Stage 4 to Stage 6, we observed FEV+ cells present at each sampled timepoint (Fig. 3.11c). To determine how transcriptionally similar the FEV+ cells found in beta cell differentiation *in vitro* were to FEV+ progenitors found in human endocrine cell development *in vivo*, we performed a Pearson's correlation analysis. Correlation analysis of FEV+ progenitors from the 12wpc_1 and 12wpc_2 datasets compared to the hESC-derived FEV+ cells revealed higher transcriptional correlation of all FEV+ clusters than compared to all FEV- clusters (Fig. 3.11d), suggesting that cells undergoing *in vitro* differentiation to a beta cell fate transit through a FEV-expressing stage similar to that found *in vivo*.

Mapping *in vitro* beta cell differentiation at single-cell resolution

We next wanted to reconstruct the lineage relationships among hESC-derived endocrine cells during *in vitro* beta cell differentiation. We first merged all CHGA+ endocrine clusters from each sampled timepoint using Seurat 3 (Fig. 3.12a, b). Given that batch correction via Seurat

3's integration method is currently not compatible with pseudotemporal ordering by Monocle 3, we utilized Monocle 3's internal batch correction method to merge hESC-derived endocrine clusters from different timepoints and subsequently performed pseudotemporal ordering. The result was one main trajectory, which we used in subsequent analyses, that began with ES4 cells, our first sampled timepoint and the designated start of our pseudotemporal ordering (Fig. 3.12c, d). As pseudotime progressed, three main endpoints of the differentiation trajectory were observed (Fig. 3.12c, d). As expected, a beta cell lineage that expressed *INS* constituted one endpoint and was primarily composed of S6D4 and S6D10 cells (Fig. 3.12e). A second endpoint was composed of poly-hormonal cells that expressed *INS, GCG*, and *SST* (Fig. 3.12e). These polyhormonal cells also were derived from the S6D4 and S6D10 timepoints (Fig. 3.12c). A third endpoint in the differentiation trajectory surprisingly resulted from a bifurcation event early in pseudotime before the acquisition of hormone identity (Fig. 3.12f). The cells at this third endpoint did not express *INS, GCG, or SST* (Fig. 3.12e). Instead, this population appeared to be mis-differentiated and expressed transcription factors such as *PHOX2A, TLX2*, and *TBX2* (Fig. 3.12f), all of which regulate differentiation and function of cells in the nervous system. These transcription factors were not expressed in a large fraction of the endocrine compartment of human fetal pancreata (Fig. 3.12g). *PHOX2A* is required for proper differentiation of neurons in the autonomic nervous system, as it is crucial for the development of neural crest-derived cells (Borghini et al., 2006; Hirsch et al., 1998; Lo et al., 1999; Pattyn et al., 1997; Tiveron et al., 1996). *TLX2* is a transcriptional target of the *PHOX2* family of transcription factors and is also required for proper development of the neural crest lineage and thus, also the enteric nervous system (Borghini et al., 2006). *TBX2* promotes anterior neural specification by suppressing FGF signaling (Cho et al., 2017). *FEV* was still expressed in this mis-differentiated lineage (Fig. 3.12e), suggesting that there may be a subset of FEV+ endocrine progenitors during *in vitro* beta cell differentiation that improperly differentiated into a neural lineage (Fig. 3.12h). The

expression of these transcription factors in this blocked cell type within our *in vitro* beta cell differentiation suggests that these cells have mis-differentiated into a neural identity.

FEV appears to be required for proper human beta cell differentiation and function

In addition to serving as a marker for endocrine progenitor stages in human endocrine cell development, we wanted to determine if *FEV* had any functional role in beta cell differentiation. In *Fev* knockout (KO) mice, glucose clearance from the blood following a glucose challenge was significantly slowed, and insulin content of beta cells was decreased (Ohta et al., 2011). Given that this study utilized a whole-body *Fev* KO, the defects in glucose homeostasis and reduction in insulin content could have been a result of a requirement of *Fev* in non-pancreas cells or *Fev* function in multiple stages in the lifetime of a beta cell. To test the requirement of *FEV* in human beta cell differentiation and function, we used our *in vitro* beta cell differentiation platform and first generated a FEV-KO hESC line through CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9-mediated genomic editing (Fig. 3.13a). The human FEV locus contains three exons, and the guide RNA (gRNA) we designed targeted the end of exon 1 (Fig. 3.13a). Wildtype hESCs were nucleofected with Cas9 and a FEV-KO gRNA, cultured for 2 passages following nucleofection to allow for recovery, and then clonally plated (Fig. 3.13b). We identified one clone that exhibited a 1-bp insertion in one *FEV* allele and a 1-bp insertion 1 in the second allele, both located at the end of exon 1 (Fig. 3.13a). Both genomic edits result in frameshift mutations that changed the entire amino acid sequence following exon 1. The DNA-binding domain normally found in exon 3 was predicted to no longer be properly translated in the new amino acid sequence following these frameshift mutations.

This FEV-KO clonal hESC line, along with a WT control clonal line, was expanded and subsequently adapted to suspension-based culture for the *in vitro* beta cell differentiation platform. Directed differentiation of both FEV-KO and WT control hESCs was then performed

(Fig. 3.13c). Ablation of *FEV* did not affect pluripotency of hESCs, as measured by OCT4 and TRA-1-60 staining (Fig. 3.13d). Ablation of FEV also did not inhibit differentiation into Stage 2 definitive endoderm as marked by *SOX17* and *FOXA2* (Fig. 3.13d), which is concordant with the lack of *FEV* expression before Stage 2. By the Stage 6 beta-like cell stage, however, *FEV* deficiency did result in a reduction of CHGA+/CPEP+ cells (21.2% in WT differentiation vs. 11.6% in FEV-KO differentiation) (Fig. 3.13d), suggesting that *FEV* is needed for proper differentiation into the hormone-expressing beta lineage.

Generation of new tools and platforms for understanding human beta cell development: Identification of FEV transcriptional targets

Our discovery that the transcription factor *FEV* is expressed in hESC-derived endocrine progenitors and beta-like cells prompted us to generate tools through which we could interrogate the function of the *FEV* gene. Through CRISPR/Cas9-mediated genomic editing, we first constructed a FEV-MYC hESC line in which a MYC epitope tag was fused to the endogenous FEV transcription factor at the C-terminus (Fig. 3.14a). This tagging would allow us to perform ChIP-seq to identify transcriptional targets of FEV (Fig. 3.14b). To engineer this line, we designed a FEV-KI (knock-in) gRNA that targeted the end of exon 3 of the *FEV* locus (Fig. 3.14a). We obtained a commercially-synthesized DNA targeting template containing 3 sequential MYC epitopes separated by small genomic spacers (termed 3xMYC) (Fig. 3.14a). This 3xMYC sequence was flanked by homology arms found around the cut site in the endogenous FEV locus (Fig. 3.14a). The 3xMYC sequence with the flanking homology arms was cloned into a pUC19 vector and transformed into competent cells in order to obtain sufficient DNA quantity for PCR amplification of the targeting template. This targeting template along with the FEV-KI gRNA and Cas9 protein were then nucleofected into wildtype hESCs, cultured for 2 passages following nucleofection to allow for recovery, and then clonally plated. The *FEV* locus of isolated clones was screened for successful knock-in of the 3xMYC by PCR

amplification of the knock-in region and Sanger sequencing the resulting PCR amplicon. This screening strategy identified a FEV-MYC hESC clonal line with one FEV allele that showed successful knock-in of the 3xMYC tag in frame with the FEV locus (Supplementary Data 20).

Given that FEV is a transcription factor that appears to be required for proper endocrine differentiation in *in vitro* beta cell differentiation, the generation of this FEV-MYC hESC line will be valuable in interrogating the mechanism through which FEV regulates proper human beta cell differentiation. ChIP-seq on FEV+ endocrine progenitors at Stage 5 of our *in vitro* differentiation can identify transcriptional targets of FEV (Fig. 3.14b), which will serve as candidate effectors of proper endocrine cell differentiation. Additionally, *FEV* is expressed in Stage 6 cells when cells of the beta lineage begin to form. In Stage 6, we found non-beta, FEV-expressing cells that appeared to be blocked in their differentiation potential (Fig. 3.14b). ChIP-seq on these blocked cells may identify transcriptional targets of FEV that may mediate improper beta cell differentiation. *Fev-KO* mouse studies demonstrated that *Fev* binds to the insulin promoter to promote *Ins* transcription (Ohta et al., 2011). ChIP-seq on sorted INS+ beta cells from Stage 6 will also identify FEV targets that regulate beta cell function and can confirm if *INS* is also a target of FEV in human beta cells.

Generation of new tools and platforms for understanding human beta cell development: Isolation of FEV-expressing cells during *in vitro* beta cell differentiation

We have generated tools with which we could isolate and characterize the FEV-expressing population during human beta cell differentiation. We constructed two *FEV* reporter hESC lines: a FEV-GFP line and a FEV-tNGFR (truncated Nerve Growth Factor Receptor) line (Fig. 3.15a). The FEV-GFP line will allow for fluorescent-based isolation of FEV-expressing cells during *in vitro* beta cell differentiation. For applications that require quicker isolations of much larger quantities of FEV-expressing cells than what fluorescent-based sorting can provide, the FEV-tNGFR line can be utilized. The tNGFR is a surface marker in which the cytoplasmic

intracellular signaling domain of the NGFR is removed and thus can be leveraged for magnetic bead-based isolation methods (Dever et al., 2016). This tNGFR enrichment strategy has already been implemented in human clinical studies for the isolation of large quantities of tNGFR-tagged cells (Bonini et al., 2003; Oliveira et al., 2015). Both the GFP and the tNGFR sequences used for these KI lines were preceded by a T2A element that would enable bicistronic translation of FEV and the reporter protein. Similar to our strategy with the FEV-MYC hESC line generation, we obtained commercially-synthesized DNA targeting templates containing either the 2A-GFP or 2A-tNGFR sequences flanked by homology arms around the cut site. Following cloning and subsequent PCR amplification of the FEV-GFP or FEV-tNGFR targeting templates, they were nucleofected into wildtype hESCs along with the same FEV-KI gRNA used for FEV-MYC line generation and Cas9 protein. The nucleofected hESCs were cultured for 2 passages to allow for recovery and then clonally plated. The *FEV* loci of isolated clones were screened for successful knock-in of the 2A-GFP or 2A-tNGFR sequences by PCR amplification of the knock-in region and Sanger sequencing the resulting PCR amplicon. This screening strategy identified both a FEV-GFP and FEV-tNGFR hESC clonal line with one *FEV* allele that showed successful knock-in of the reporter in frame with the *FEV* locus (Supplementary Data 21 and 22).

Purification of FEV-expressing cells at defined stages of the *in vitro* beta cell differentiation process will enable us to understand the differences among FEV-expressing populations at each differentiation stage. Specifically, purifying FEV+ endocrine progenitors at stage 5 of our differentiation program will permit small molecule screens to identify compounds that can either induce progenitor expansion prior to beta cell lineage commitment or enhance differentiation toward the beta lineage (Fig. 3.15b). Similarly, given that we found FEV-expressing cells that appeared to be mis-differentiated in Stage 6, we can also harness the utility of these reporter lines to isolate these mis-differentiated cells and screen for compounds that can correct their differentiation into the beta lineage or block their emergence altogether (Fig. 3.15c). Finally, given that beta cells in human pancreatic development transit through a

FEV-expressing precursor stage, we can test whether enriching for FEV+ endocrine progenitor-stage cells during Stage 5 will yield higher efficiencies of beta cell differentiation at Stage 6. Given that large quantities of cells would be needed, utilizing our FEV-tNGFR line would allow us to isolate enough FEV+ endocrine progenitors in Stage 5 that we can re-aggregate into clusters for differentiation into Stage 6. Thus, the ability to isolate FEV-expressing cells throughout *in vitro* beta cell differentiation will refine our understanding of the cellular heterogeneity that emerges during this directed differentiation process.

Generation of new tools and platforms for understanding human beta cell development: Validation of novel candidate regulators of beta cell lineage allocation and function

Our *in vivo* and *in vitro* single-cell RNA-sequencing analyses have resulted in the identification of candidate beta cell lineage regulators that require validation. In order to functionally validate these candidate regulators, we developed a flexible platform on which we can test whether these genes regulate beta cell lineage allocation (Fig. 3.16a). In this platform, we design different gRNAs that will introduce a frameshift mutation within the genomic locus of each candidate regulator we wish to functionally test. Given that these genes are candidate regulators of endocrine lineage allocation, we utilize the endocrine progenitor-stage (Stage 5) clusters from our *in vitro* beta cell differentiation and dissociate them into a single-cell suspension. We then nucleofect endocrine progenitor-stage cells with Cas9 and a specific gRNA against the candidate regulator of interest. Since editing is not 100% efficient, nucleofection of these gRNAs will lead to a knockdown, not a full knock-out, of the candidate regulator of interest. Following nucleofection, we reaggregate the endocrine progenitor-stage cells into clusters and proceed with the directed differentiation towards the beta lineage. This platform leveraging both *in vitro* beta cell differentiation and temporally controlled CRISPR/Cas9-mediated genomic editing provides a versatile solution to functionally validate the candidate regulators identified through *in silico* methods.

Discussion

Redefining the NGN3+ endocrine progenitor population in human pancreatic development

In human endocrine cell development, NGN3 has long been thought to mark the endocrine progenitor population, given the function of *Ngn3* in mouse pancreatic development. Indeed, NGN3 is required for endocrine cell differentiation in human endocrine cell development, as inactivating mutations of *NGN3* lead to neonatal diabetes (Pinney et al., 2011; Wang et al., 2006). Beta cell mass is suspected to be reduced, not absent, in human cases of inactivating *NGN3* mutations given that C-peptide is detected in the blood, albeit at low levels (Pinney et al., 2011). This is in contrast to mouse development, in which *NGN3* ablation halts beta cell generation altogether (Gradwohl et al., 2000). Through our study of human endocrine cell development, NGN3 did not appear to be the most robust marker of the endocrine progenitor population common to hormone-expressing lineages, such as the alpha and beta lineages, in our 12wpc_1 human fetal pancreas dataset. Other markers that appeared to more faithfully label this common endocrine progenitor population included *EMC10, SOX4, HES6*, and *KRT19*. *CTD-2545M3.8* also emerged from our differential gene expression analysis as a marker specific to this common endocrine progenitor population but awaits functional characterization. It is possible that *NGN3* was already been downregulated in this common endocrine progenitor population and that we did not capture enough *NGN3*-expressing progenitors for UMAP-based clustering to categorize them in their own distinct cluster. If this is true, this may imply that the period between *NGN3* downregulation and differentiation into a hormone-expressing lineage is proportionally much longer in human development than in mouse. Time-lapsed imaging of live, human fetal pancreatic explants will provide a more refined profile of NGN3 expression dynamics in human pancreatic development.

In murine endocrine development, Ngn3+ endocrine progenitors give rise to all five hormone-expressing lineages of the pancreas (Gradwohl et al., 2000; Heller et al., 2005).

However, in the human fetal pancreas, the *NGN3*-expressing common endocrine progenitor population appeared to only give rise to alpha and beta lineages. In our pseudotemporal ordering analysis, there was no trajectory that connected *NGN3*-expressing progenitors to the *SST*-expressing delta lineage or the *GHRL*-expressing epsilon lineages. We did not observe a distinct *PPY*-expressing gamma population, as all *PPY*-expressing cells also expressed *GCG* and thus were annotated as alpha cells. This differentiation potential of the *NGN3*-expressing common progenitors that we identified in human pancreatic development could be the result of not obtaining enough cells for pseudotemporal ordering analysis, which relies on similarities in gene expression profiles to order cells along a differentiation trajectory. This result also was derived from one timepoint in fetal development, and the development of improved batch correction methods for pseudotemporal ordering may allow us to combine multiple timepoints together for more powerful analysis. However, if this discrepancy in the differentiation potential reflects true lineage relationships of NGN3-expressing endocrine progenitors in the developing human pancreas, this would depart from the dogma established by findings of the lineage potential of Ngn3+ progenitors in murine pancreatic development. Future work clarifying the hormone-expressing lineages derived from NGN3+ endocrine progenitors in human fetal development is needed.

Identification of novel pre-alpha and pre-beta cell stages in human pancreatic development

Mapping endocrine cell development at a higher resolution using single-cell RNA-sequencing can be leveraged for developing new methods to generate endocrine cell types more efficiently from stem cell sources. In this work, we have identified novel pre-alpha and pre-beta progenitor stages that provide increased resolution regarding the steps required to differentiate into alpha or beta lineages in human pancreatic development. While a handful of other single-cell RNA-sequencing studies have identified additional endocrine progenitor stages

in mouse pancreatic development, to our knowledge, no published work has identified novel endocrine progenitor stages in human pancreatic development using single-cell RNA-sequencing. Our work in human fetal pancreatic development thus offers novel endocrine progenitor stages onto which we can compare and contrast the biological relevance of these murine progenitor stages to those of human.

The advent of single-cell RNA-sequencing has led to the discovery of several novel endocrine progenitor stages in mouse pancreatic development. One example is our own work, in which we discovered a novel intermediate endocrine progenitor population defined by high *Fev* expression (Byrnes et al., 2018). *Fev* expression has also been identified in endocrine progenitor populations reported by several other single-cell RNA-sequencing studies of murine pancreatic development (Krentz et al., 2018; Scavuzzo et al., 2018), confirming the reproducibility of our finding. This Fev+ endocrine progenitor is derived from an Ngn3+ population, and differentiated endocrine lineages in the murine pancreas transit through a Fev-expressing cell stage (Byrnes et al., 2018).

Within the Fev+ progenitor population, we also identified cells that appeared to be pre-specified towards an alpha or beta cell fate (Byrnes et al., 2018). This is analogous to human pancreatic development in which we not only identified endocrine progenitors that expressed *FEV* but also observed that these FEV-expressing progenitors appeared to be already lineage-specified towards an alpha or beta cell fate. Our *in silico* reconstruction of endocrine lineage relationships suggested that endocrine cell fate decisions in progenitors occurs at the *Fev*/*FEV*-expressing cell stage in both mouse and human.

Beyond this *Fev*-expressing endocrine progenitor stage, there are additional endocrine progenitor stages that have been identified in murine development. In particular, four distinct endocrine progenitor stages (termed EP1-4) have been proposed in mouse endocrine cell development (Yu et al., 2019). Expression of *Ngn3*, the canonical pro-endocrine lineage marker in pancreatic development, increased in EP1, peaked in EP2, decreased in EP3, and was not

observed in EP4. Expression of *Fev* was found in EP3 and EP4 stages only (Yu et al., 2019), which is concordant with *Fev* being downstream of *Ngn3* (Byrnes et al., 2018; Miyatsuka et al., 2014). Interestingly, many of the differentially expressed genes in each EP stage were also identified as top differentially expressed genes in either our human endocrine progenitor clusters or during pseudotemporal ordering. Specifically, *Krt19* and *Gadd45a*, two genes that defined a human common endocrine progenitor stage in our dataset of 12wpc_1 human fetal pancreas, were found to be differentially expressed in EP2. Several candidate beta lineage regulators in human fetal development were also found in EP1 (*Arid5b*), EP3 (*Ahi1*), and EP4 (*Rbp4, Peg10, Acvr1c, Sez6l2*) (Yu et al., 2019). Similarly, several candidate alpha lineage regulators in human fetal development were found in EP3 and EP4 (A*rx, Irx2, Fam46a, Slc30a8, Slc7a2, Slc7a8, Cryba2, St18,* and *Alcam*) (Yu et al., 2019). Thus, these EP stages found in murine endocrine development appear to also have relevance to endocrine progenitor stages found in human fetal pancreatic development, and we believe it is important to confirm these murine progenitor stages in human pancreatic development.

Transcriptional mechanisms underlying fate decisions are shared across tissues

Our single-cell RNA-sequencing analysis of human endocrine lineage allocation identified many candidate regulators previously identified and studied in the nervous system. Despite their derivation from different germ layers, both the pancreatic endocrine and neural lineages employ many of the same transcription factors that regulate their own development, including *Ngn3, NeuroD1, Nkx2.2, Nkx6.1, Pax* family of transcriptional regulators*,* and *Fev* (Blake and Ziman, 2014; Churchill et al., 2017; Gradwohl et al., 2000; Hendricks et al., 1999; Mastracci et al., 2013; Napolitano et al., 2015; Ohta et al., 2011; Pataskar et al., 2016; Prakash et al., 2009; Qi et al., 2001; Schaffer et al., 2010; Simon-Areces et al., 2010; St-Onge et al., 1997) . These transcriptional similarities have an evolutionary basis, as the main source of insulin in invertebrates is in neurons (Wong et al., 2014). Thus, from an evolutionary

perspective, it is not surprising that additional genes previously identified to be required for proper nervous system development and function are also implicated in pancreatic endocrine development and, more specifically, lineage allocation. As functional validation is performed of these candidate regulators of pancreatic endocrine lineage allocation, key insights and similarities may be drawn from understanding the transcriptional hierarchies that govern both pancreatic endocrine and neural developmental programs.

The development of enteroendocrine cells (EEs) in the intestine also shares striking similarity to pancreatic endocrine cell development. Proper differentiation of EEs in the intestine during development requires transcription factors also critical for pancreatic endocrine cell differentiation, including *Ngn3* (Jenny et al., 2002; López-Díaz et al., 2007; Schonhoff et al., 2004), *Nkx2.2* (Gross et al., 2016), *Isl1* (Terry et al., 2014), *NeuroD1* (Naya et al., 1997), *Pax4* (Beucher et al., 2012a). As in pancreatic endocrine cell development, the EE lineage comprises multiple hormone-expressing cell types that are derived from a common progenitor cell defined by *Ngn3* (Jenny et al., 2002). Recent work applying single-cell RNA-sequencing to murine EE development uncovered novel markers and lineage-specific regulators of the multiple EE lineages (Gehart et al., 2019), and many of these genes overlapped with the markers and candidate transcriptional regulators that we identified in mouse and human endocrine cell development and lineage allocation. Ngn3+ EE progenitors differentially express *Sox4, Tox3*, and *Gadd45a* (Gehart et al., 2019), all of which were also defining markers of our common endocrine progenitor in human endocrine cell development. Known hormone-specific lineage regulators in pancreatic endocrine cell development, such as *Arx, Pax6,* and *Isl1*, were also identified as EE-specific lineage regulators (Gehart et al., 2019). Interestingly, a number of novel candidate lineage regulators that we identified in mouse and human endocrine lineage allocation were also found to be lineage-specific regulators of the different EE lineages (Gehart et al., 2019). These include *Nr4a2, Smarca1, Peg3, Id1, S100a1,* and *Klf4* (Gehart et al., 2019). Thus, given these transcriptional similarities between the hormone-expressing lineages of the

pancreas and intestine, future work focused on deciphering additional lineage-specific regulators of the multiple hormone-expressing EE lineages may provide additional insight into transcriptional regulation of pancreatic endocrine cell fate selection.

Timing of endocrine lineage fate decisions

The timing of endocrine lineage fate commitment in pancreatic development is not fully understood. The Ngn3+ endocrine progenitor stage has long been regarded as the master stage prior to endocrine cell differentiation, but single-cell RNA-sequencing studies of pancreatic development, including our own, have identified additional progenitor stages that arise between initial *Ngn3* expression and acquisition of differentiated cell identity. This increased resolution of endocrine cell differentiation has provided us with new cell stages that we can interrogate for determining when endocrine lineage decisions are made. From both our mouse and human studies of endocrine cell development, Fev/FEV-expressing endocrine progenitors were already specified towards an alpha or beta cell fate (Byrnes et al., 2018). This heterogeneity in Fev/FEV-expressing progenitors suggests that endocrine lineage specification occurs at this Fev/FEV-expressing progenitor stage or prior. With the advent of single-cell ATAC-seq, future work can look beyond the transcriptome and into the epigenome to determine if certain subsets of endocrine progenitors are primed to express alpha or beta-specific lineage genes or are pre-committed to a specific hormone lineage. This fate commitment could also possibly be made in Pdx1+ pancreatic progenitors prior to the expression of pro-endocrine genes, such as *Ngn3*, altogether, although our single-cell RNA-sequencing combined with pseudotemporal ordering identified endocrine progenitor populations that appeared to be fated towards one specific endocrine lineage. Future transcriptomic and epigenomic profiling of stages preceding endocrine differentiation will help determine the exact stage at which endocrine lineage determination occurs.

The timing of endocrine fate decisions can also be regulated by extrinsic signals derived from the surrounding microenvironment. In murine development, the developmental time at which Ngn3+ progenitors form corresponds to their ultimate hormone lineage selection (Johansson et al., 2007). The competence window for alpha differentiation arises the earliest in murine pancreatic development, resulting in alpha cells being the first emerging endocrine lineage, followed by beta and gamma cells, and then lastly followed by delta cells (Johansson et al., 2007). In contrast, in human pancreatic development, the beta lineage is the earliest endocrine cell type to be detected (at 6wpc), followed by alpha cells (at 8-9wpc), delta cells (10wpc), and gamma cells (at 17wpc) (Jeon et al., 2009; Piper et al., 2004). The differences in timing of emergence of endocrine lineages between mouse and human could be a direct result of the changing microenvironment during development that can be providing dynamic cues that promote one endocrine lineage over the other. From murine studies, we know that several compartments of the microenvironment influence pancreatic development, including vasculature, nerves, and mesenchyme (Borden et al., 2013; Golosow and Grobstein, 1962; Landsman et al., 2011; Magenheim et al., 2011; Reinert et al., 2013). However, the cellular composition of each microenvironment compartment can widely differ between that of mouse and human. From our single-cell profiling of human fetal pancreas, we identified several populations of endothelial cells whose transcriptional expression profiles changed throughout the course of development. These changes may influence the competency of endocrine progenitors to differentiate into distinct hormone lineages, either through secreted signaling molecules or direct interactions. Our single-cell profiling in both mouse and human pancreatic development also reflects different mesenchymal and nerve populations whose dynamics may regulate endocrine differentiation at distinct periods in development.

FEV in human endocrine cell differentiation and function

Fev is a transcription factor best known for its role in serotonergic neuron differentiation and function (Haugas et al., 2016; Wyler et al., 2016), and more recently, *Fev* has also been implicated as a regulator of the differentiation of multiple hormone-expressing enteroendocrine lineages in the intestine (Gehart et al., 2019). Given that hormone-producing cells in the brain and hormone-producing cells in the pancreas are evolutionarily related, *Fev* function was previously investigated in the islet lineages of the pancreas, where it was found to regulate insulin production in adult mouse beta cells (Ohta et al., 2011). *Fev-KO* mice exhibited defects in glucose clearance from the blood following a glucose tolerance test, and this defect was attributed to decreased insulin content in beta cells (Ohta et al., 2011). Following ChIP-seq studies in a beta cell cell line, Fev was identified to bind directly to the *Insulin* promoter, thus regulating its transcription and resulting in less *Insulin* transcription with the loss of *Fev* in beta cell (Ohta et al., 2011). Beta cell mass in *Fev-KO* mice was not decreased, indicative of decreased insulin production across the entire pancreas in *Fev-KO* mice as opposed to fewer beta cells present in the adult pancreas (Ohta et al., 2011). *Fev* ablation also did not result in obvious defects in the differentiation of alpha, beta, delta, epsilon, and gamma cells during embryonic development (Ohta et al., 2011), although a robust quantification to assess possible skews towards specific endocrine lineages with the loss of *Fev* should be performed.

Our work investigating the role of *FEV* in human endocrine cell differentiation and function highlights potential differences between *Fev/FEV* in mouse versus human. In both mouse and human endocrine cell development, *Fev/FEV* was expressed in an intermediate progenitor stage that followed initial NGN3 expression and preceded hormone acquisition (Byrnes et al., 2018). *FEV* was also expressed in endocrine progenitor stage cells during *in vitro* beta cell differentiation from hESCs, which was concordant with *FEV* expression in endocrine progenitors in human fetal pancreatic development. Notably, although *Fev-KO* mice do not exhibit obvious differentiation defects in the islet lineages during development, we did observe a

reduction in the differentiation into CHGA+/CPEP+ beta cells in our *in vitro* beta cell differentiation model. This suggests that *FEV* may be required for human beta cell differentiation and is dispensable for mouse beta cell differentiation. Quantification of the beta lineage during pancreatic development in *Fev-KO* mice will confirm whether this discrepancy between mouse and human is indeed true.

Notable differences were also observed between *Fev/FEV* in mouse and human differentiated endocrine cells. While *Fev* expression persists in the alpha and beta lineages during mouse pancreatic development, *FEV* expression was downregulated in beta cells and only maintained in the alpha lineage in human pancreatic development. Single-cell RNA-sequencing of adult human islets has indicated that *FEV* is expressed in alpha cells and not beta cells (Segerstolpe et al., 2016). This is in contrast to our *in vitro* beta cell differentiation system, in which beta cells appeared to maintain *FEV* expression following differentiation of the FEV+ endocrine progenitor stage. In mouse beta cells, *FEV* binds to the insulin promoter to regulate *Insulin* transcription and thus insulin production. Given that *FEV* turns off in differentiated human beta cells *in vivo*, it is possible that *FEV* is either not needed for beta cell function or *FEV* inhibits beta cell function. In our *in vitro* beta cell differentiation, we actually observed a subset of INS+ beta cells that did not express *FEV*, whereas another subset of INS+ beta cells did express FEV. The INS+/FEV- hESC-derived beta cells may correspond to bona fide beta cells found *in vivo*, and the INS+/FEV+ hESC-derived beta cells may either be mis-differentiated or on their way towards a FEV- state. This discrepancy in *Fev/FEV* expression during endocrine cell differentiation is just one example that highlights the need for validating findings from mouse studies in human studies.

Identification of FEV transcriptional targets will illuminate a clearer picture of its function. In the human beta cell lineage, loss of *FEV* coincided with a reduction in beta cell differentiation. Given that *FEV* was expressed in pre-beta progenitors *in vivo* and hESC-derived endocrine progenitor stage cells, it is possible that *FEV* serves as a key transcriptional regulator for

differentiation from a progenitor to a beta cell. Using our FEV-MYC hESC line during *in vitro* beta cell differentiation and performing ChIP-seq on FEV+ endocrine progenitor stage cells can identify transcriptional targets that may mediate the transition from a pre-beta progenitor to a differentiated beta cell. By acquiring a transcriptional map of FEV transcription factor activity, we may be able to modify our current *in vitro* beta cell differentiation protocol to one that promotes the expression of key FEV-regulated transcriptional circuits that promote beta cell differentiation from endocrine progenitors.

Identification of transcriptional targets in hESC-derived FEV+ beta cells will also illuminate the function of *FEV* in differentiated beta cells. Given that *FEV* was expressed in pre-alpha progenitors *in vivo* and was maintained in differentiated alpha cells, identifying transcriptional targets during alpha cell differentiation and function and how they differ from those identified in the beta lineage will augment our understanding of endocrine cell specification. Finally, ChIP-seq on endocrine lineages in the human fetal pancreas across multiple timepoints can also be performed once a suitable FEV antibody for ChIP-seq is generated.

Suppressing the formation of hESC-derived blocked endocrine progenitors

Tremendous effort has been devoted to determining the molecular cues that will make derivation of the beta lineage from hESCs more efficient. Given that we uncovered a hESC-derived *FEV*-expressing population that appeared to be mis-differentiated during *in vitro* beta cell differentiation, we can begin to uncover the molecular circuitry that pushes hESC-derived cells down this mis-directed lineage. The top differentially-expressed gene of this blocked, *FEV*-expressing cell population was the transcription factor *PHOX2A*. Interestingly, *PHOX2A* was also previously reported to mark a non-endocrine population that emerged in *in vitro* beta cell differentiation but was not described as mis-directed in differentiation potential (Veres et al., 2019). *PHOX2A* is a pro-neural homeodomain transcription factor and a key regulator of neural

progenitor differentiation into noradrenergic neurons of the central nervous system (CNS) and the peripheral nervous system (PNS) (Lo et al., 1998; Morin et al., 1997). Noradrenergic neurons are characterized by synthesis and storage of catecholamines, including norepinephrine, which serve as neurotransmitters (Hayashida and Eisenach, 2018). In this differentiation process, BMP2 and cyclic AMP (cAMP) signaling synergistically induce noradrenergic neuron differentiation through *Phox2a* transcription and *Phox2a* activation (Benjanirut et al., 2006; Chen et al., 2005; Paris et al., 2006). Given the expression of *PHOX2A* specifically in cells occupying the mis-differentiated trajectory in our pseudotemporal ordering analysis, it is possible that the PDX1+/NKX6.1+ pancreatic progenitors in our *in vitro* beta cell differentiation process have mis-differentiated towards this PHOX2A+ noradrenergic neural lineage.

Inhibition of BMP2 and cAMP signaling represents possible avenues through which we can prevent *in vitro* beta cell differentiation from entering down this mis-directed differentiation path that resembled the noradrenergic neural lineage. Interestingly, a recent iteration of our *in vitro* beta cell differentiation process was published adding LDN193189, a BMP inhibitor, to endocrine progenitor stage cells (Veres et al., 2019). LDN193189 is typically used at the beginning of the pancreatic progenitor stage to inhibit BMP signaling but is removed from the cocktail of exogenous growth factors following this stage (Pagliuca et al., 2014). This newer iteration of *in vitro* beta cell differentiation resulted in the generation of beta cells that were functional and maintained stimulation indices for up to four weeks in culture, but it also contained modifications to other stages, including removal of various growth factors from Stage 3 pancreatic progenitors (SANT1, LDN193189, and PdBU), Stage 4 pancreatic progenitors (SANT1 and RA and Rock inhibitor), and Stage 6 beta-like cells (Alk5i and T3 were removed). It is possible that the addition of LDN193189 to Stage 5 endocrine progenitors inhibited the formation of cells blocked in differentiation potential and generated beta cells with a more stable functional phenotype, although future work will need to test this possibility in a more controlled

manner. Given that these mis-differentiated, noradrenergic-like cells emerge in Stage 5 and persist into Stage 6, addition of BMP inhibitors such as LDN193189 to Stage 5 and 6 cultures can potentially block this mis-differentiation into PHOX2A+ noradrenergic-like cells and increase the efficiency at which hESC-derived endocrine progenitors differentiate into the beta lineage. Addition of cAMP signaling inhibitors, such as H89 and okadaic acid, that have been used to block PHOX2A-mediated differentiation of neural progenitors into noradrenergic lineages can also be added in order to potentially suppress mis-differentiation of endocrine progenitors during *in vitro* beta cell differentiation (Chen et al., 2005; Shin et al., 2009).

Reported enterochromaffin cells in *in vitro* beta cell differentiation

Currently, *in vitro* beta cell differentiation does not result in 100% purity of beta cells, and there are other cell types that arise during the directed differentiation of hESCs to the beta lineage. Recently, a population deemed enterochromaffin cells (ECs) has been described to arise in *in vitro* beta cell differentiation (Veres et al., 2019). ECs reside along the epithelial lining of the intestine and are the most abundant cell type among the enteroendocrine cells found in the intestine (Lund et al., 2018). The main functions of ECs are to regulate intestinal motility required for digestion and modulate the activity of the enteric nervous system through the production and secretion of the neurotransmitter serotonin. Although ECs make up less than 1% of the total intestinal epithelium, they produce more than 90% of the body's serotonin (Gershon, 2013; Mawe and Hoffman, 2013). Unlike neurons, ECs utilize tryptophan hydroxylase 1 (TPH1) and not TPH2 to synthesize serotonin, and instead of employing small neurosecretory vesicles, ECs store serotonin in large dense core vesicles (LDCVs) with the help of CHGA and CHGAB (Côté et al., 2003; Machado et al., 2010; Walther and Bader, 2003). Thus, ECs resemble lineages of both the nervous system and hormone-secreting pancreatic islets.

ECs are defined by the expression of markers that also are expressed by both serotonergic neurons and pancreatic endocrine cells. These markers include *Fev, Lmx1a,*

Lmx1b, and *Tph1* (Ding et al., 2003; Kiyasova and Gaspar, 2011; Liu et al., 2010; Maurer et al., 2004; Ohta et al., 2011; Wyler et al., 2016; Zhang et al., 2017). Proper differentiation of ECs in the intestine during development also requires transcription factors critical for pancreatic endocrine cell differentiation, including *Ngn3* (Jenny et al., 2002; López-Díaz et al., 2007; Schonhoff et al., 2004), *Nkx2.2* (Gross et al., 2016), *Isl1* (Terry et al., 2014), *NeuroD1* (Naya et al., 1997), *Pax4* (Beucher et al., 2012a). Interestingly, *Fev* is also expressed by ECs but is not required for EC differentiation in mice (Wang et al., 2010b). Given the striking similarity in gene expression profiles of ECs and EC differentiation to those of pancreatic endocrine cells, it is not surprising to observe ECs generated in *in vitro* beta cell differentiation. The mis-differentiation of hESC-derived endocrine progenitors towards similar lineages, such as that of the EC, is not surprising, given that *in vitro* beta cell differentiation is not 100% efficient. It is likely that the ECs observed during *in vitro* beta cell differentiation represent another mis-differentiation process, similar to that observed with mis-differentiated PHOX2A+ cells. Given that these hESC-derived ECs also express *FEV*, these ECs could be mistaken for the FEV+ endocrine progenitors we identified in human pancreatic development. However, this is not the case, given that a separate population of FEV+ hESC-derived endocrine progenitors is found and that give rise to the beta lineage in not only our *in vitro* differentiation process but also in that reported by others (Veres et al., 2019). It is likely that these FEV+ ECs are derived from the same hESC-derived FEV+ endocrine progenitors that give rise to beta cells. The molecular cues that drive differentiation into the EC lineage over the beta lineage will be important to dissect in order to improve the outcomes of *in vitro* beta cell differentiation. The mis-differentiation into the EC lineage, coupled with the emergence of mis-differentiated PHOX2A+ cells, highlights the need for a more refined understanding of the transcriptional machinery that pushes pancreatic and endocrine progenitors toward the beta lineage.

Materials and Methods

Human tissue procurement, isolation, and processing

Human fetal pancreata were harvested from post-mortem fetuses with approval from the ethical committee of UCSF. Tissue was obtained through two sources: the University of Washington Birth Defects Research Laboratory (12wpc_1 and 15.5wpc samples) and Advanced Bioscience Resources, Inc (12wpc_2 and 16wpc samples). Tissue was harvested at respective clinics and shipped overnight on ice in either 1X PBS (samples from University of Washington Birth Defects Research Laboratory) or RPMI media (samples from Advanced Bioscience Resources, Inc). Following delivery to our laboratory, tissue was washed once with 1X PBS, minced with a sterile scalpel, and dissociated in Liberase TM and 0.1 mg/mL DNase in 1X HBSS for 30-55 minutes in a 37°C Thermomixer programmed to shake at 1000rpm. Dissociation was quenched with 5mM EDTA and 10% FBS in 1X HBSS. Cell suspension was filtered through a 30um cell strainer. Red blood cells (RBCs) were removed from the cell suspension using immunomagnetic negative selection with STEMCELL Technologies' EasySep RBC Depletion Reagent (cat. no. 18170). Following RBC depletion, cells were counted and loaded onto the 10X Chromium Platform for single-cell RNA-sequencing.

Adult human islets were isolated from cadaveric donor tissue by the UCSF Islet Production Core with approval from the UCSF ethical committee. Consented cadaver donor pancreata were provided by the nationally recognized organization UNOS via local organ procurement agencies. The identifiers were maintained at the source only, and the investigators received de-identified specimens.

Informed consent was obtained for all human (fetal and adult) tissue collection, and protocols were approved by the Human Research Protection Program Committee on Human Research of UCSF.

Embryonic stem cell culture and differentiation

The hESC line HUES8 was obtained from Harvard University and used for the generation of hESC-derived beta-like cells (BLCs). Pluripotent HUES8 cells were maintained as spherical clusters in suspension in mTeSR-1 (StemCell Technologies) in 500 mL spinner flasks (Corning, VWR) on a magnetic stir plate (Dura-Mag) within a 37°C incubator at 5% CO_2, 100% humidity, and a rotation rate of 70 rpm. Cells were screened for mycoplasma contamination using the MycoProbe Mycoplasma Detection Kit (R&D Systems), according to the manufacturer's instructions.

BLCs were generated as previously described (Pagliuca et al., 2014), with additional modifications (Millman et al., 2016). In brief, HUES8 cells were seeded into a spinner flask at a concentration of $8e^5$ cells/mL in mTeSR1 media with 10 μM Rock inhibitor Y-27632 (STEMCELL Technologies) to allow formation of spherical clusters. Differentiation was initiated 72 h later. Differentiation was achieved in a step-wise fashion using the following growth factors and/or small molecules: definitive endoderm (Stage 1) (1 day of 100 ng/mL Activin A (R&D Systems) and 14 μg/mL of CHIR99021 (Stemgent); 2 days of 100 ng/mL Activin A); gut tube endoderm (Stage 2) (3 days of 50 ng/mL KGF (Peprotech)); early pancreatic progenitors (Stage 3) (1 day of 200 nM LDN193189 (Fisher Scientific), 50 ng/mL KGF, 0.25 μM SANT-1 (Sigma), 2 μM Retinoic Acid (Sigma), 500 nM PdbU (EMD Biosciences), and 10 μM Rock inhibitor Y-27632 (STEMCELL Technologies); 1 day of 50 ng/mL KGF, 0.25 μM SANT-1, 2 μM Retinoic Acid, 500 nM PdbU); later pancreatic progenitors (Stage 4) (5 days of 50 ng/mL KGF, 0.25 μM SANT-1, 0.1 μM Retinoic Acid, and 10 μM Rock inhibitor Y-27632); endocrine progenitors (Stage 5) (4 days of 0.25 μM SANT-1, 0.1 μM Retinoic Acid, 1 μM XXI (EMD Millipore), 10 μM Alk5i (Axxora),

1 μM T3 (EMD Biosciences), 20 ng/mL Betacellulin (Fisher Scientific); 3 days of 25 nM Retinoic Acid, 1 μM XXI, 10 μM Alk5i, 1 μM T3, 20 ng/mL Betacellulin); BLCs (Stage 6) (6-11 days of 10 μM Alk5i; 1 μM T3). Successful differentiation was assessed at the completion of Stages 1, 3, 4, 5, and 6 via immunofluorescence or FACS for stage-specific marker genes. hESC-derived cells used for single-cell RNA-sequencing were taken at ES4 (End of Stage 4), S5D4 (Stage 5, Day 4), S5D7, S6D4, and S6D10. Cells for single-cell RNA-sequencing were dissociated with Accumax for 15-25 mins in a 37°C water bath. The dissociated cell suspension was neutralized with stage-specific media and filtered through a 37um filter. Cells were counted and then loaded onto the 10X Chromium Platform for single-cell RNA-sequencing.

***In situ* hybridization and immunofluorescence of hESC-derived clusters**

hESC-derived cell clusters were fixed in 4% PFA in 1X PBS for 15 mins at RT. Fixed clusters were washed with 1X PBS and cryoprotected overnight at 4°C in 30% sucrose. Clusters were then embedded in OCT, and sections measuring 8 μm sections were cut.

In situ hybridization was performed on 8 μm sections using RNAscope technology (Advanced Cell Diagnostics) according to the manufacturer's instructions. An *in situ* probe against human *FEV* (cat. no. 471421-C3) was used in combination with the RNAscope Multiplex Fluorescent Reagent Kit v2 for target detection. Following signal amplification of the target probes, sections were washed in 1X PBS three times and blocked in 5% normal donkey serum (NDS, Rockland Immunochemicals) in 0.1% Triton X-100 in PBS for 1 hr at RT. Tissue sections were then stained with a primary antibody against PDX1 (1:100, R&D Systems). The next day, sections were washed three times in 0.1% Tween 20 in 1X PBS and then incubated with species-specific Alexa Fluor 488-secondary antibodies (1:500, Jackson ImmunoResearch) and DAPI in 5% NDS in 0.2% PBT for 1 hr at RT. Sections were washed three times in 0.1% Tween 20 in 1X PBS,

rinsed in 1X PBS, and then mounted in ProLong Gold Mounting Medium. Slides were stored at 4 °C.

Images were captured on a Leica confocal laser scanning SP8 microscope. Maximum intensity z-projections were then prepared using ImageJ, where brightness, contrast, and pseudo-coloring adjustments were applied equally across all images in a given series.

Quantitative RT-PCR

hESC-derived cells at various stages of directed differentiation were collected in Trizol, and RNA was extracted with the Direct-zol RNA Miniprep kit (Zymo Research). Adult human islets were also processed this same manner for RNA extraction. Reverse transcription was performed with the Superscript IV First-Strand Synthesis System (Thermo Fisher Scientific, cat. no. 18091050) using Oligo d(T) primers and random hexamers. RT-PCR was run on an ABI Real-Time PCR System (Applied Biosystems, 384-well format) with Taqman probes for *FEV* (assay ID: Hs00232733_m1) and *GAPDH* (assay ID: Hs02758991_g1) in triplicate. Data were normalized to *GAPDH*.

Single-cell capture and sequencing

To capture individual cells, we utilized the Chromium Single Cell 3' Reagent Version 3 Kit (10X Genomics) (Zheng et al., 2017). Only our 15.5wpc sample was processed with the Chromium Single Cell 3' Reagent Version 2 Kit. For all samples, 25,000 cells were loaded onto one or two wells of the 10X chip to produce Gel Bead-in-Emulsions (GEMs). GEMs underwent reverse transcription to barcode RNA before cleanup and cDNA amplification. Libraries were prepared with the Chromium Single Cell 3' Reagent Kit. Each sample was sequenced on the NovaSeq (Illumina) in Rapid Run Mode with paired-end sequencing parameters: Read1, 98 cycles; Index1, 14 cycles; Index2, 8 cycles; and Read2, 10 cycles.

Single-cell analysis

We utilized CellRanger v3.0.2 software for all single-cell RNA-sequencing datasets with default settings for de-multiplexing, aligning reads to the human genome (10X Genomics pre-built hg38 reference genome) with STAR (Dobin et al., 2012) and counting unique molecular identifiers (UMIs) to build transcriptomic profiles of individual cells. Gene-barcode matrices were analyzed with the R package Seurat v3.0.1 (Stuart et al., 2019). We first performed a filtering step, retaining only the cells that expressed a minimum and maximum number of genes and did not exceed a specified percentage of reads that map to the mitochondrial genome. The following quality control metrics for each dataset are outlined in Table 3.1.

Table 3.1. Quality control metrics for human single-cell sequencing analyses. Sample name is listed along with the minimum and maximum number of genes and maximum percentage of mitochondrial genes used for quality control thresholds.

Sample Name	Minimum number of genes	Maximum number of genes	Maximum percentage of mitochondrial genes
12wpc_1	200	4,000	15
12wpc_2	200	5,000	15
15.5wpc	200	4,000	7.5
16wpc	200	6,000	15
ES4	200	6,000	10
S5D4	200	6,000	12.5
S5D7	200	6,000	15
S6D4	200	6,000	15
S6D10	200	6,000	15

Data were then normalized with the Seurat3 function NormalizeData with default settings. This employs a global-scaling normalization that normalizes gene expression measurements for

each cell by the total expression. Genes that exhibit high cell-to-cell variation were then identified using FindVariableFeatures. The highly variable genes from this analysis were then used in downstream analysis to highlight biological signal from background noise in single-cell datasets. Data then underwent linear transformation ("scaling"), which is required prior to dimensional reduction with PCA, and this scaling was done with ScaleData. PCA (Principal Component Analysis) was performed on the scaled data with RunPCA. Significant PCs (principal components) were determined with ElbowPlot, which plots principal components based on the percentage of variance exhibited by each one. These significant PCs were utilized in Seurat3's graph-based clustering algorithms, FindNeighbors and FindClusters. The resolution parameter of FindClusters was adjusted to vary the number of clusters found by the algorithm. Clusters were visualized by UMAP with Seurat3's RunUMAP and DimPlot functions. Differentially expressed genes were determined with the FindAllMarkers function. Seurat3's VlnPlot, DotPlot, and FeaturePlot functions were used to visualize of expression of genes of interest across cells and clusters.

Sub-clustering and merging datasets

Sub-clustering was performed by isolating clusters of interest with the Seurat3 function Subset and reanalyzing as outlined above (finding variable genes, scaling data, and identification of significant PCs). Cells were classified as endocrine based on the expression of *Chromogranin A* (*CHGA*).

Merging of all human fetal datasets was accomplished with Seurat3's Integration workflow. This integration workflow in Seurat3 identifies "anchors" across disparate single-cell datasets in order to construct harmonized references for better merging of the data and minimization of batch effect (Stuart et al., 2019). In the integration workflow, all datasets were merged into a single Seurat object and processed to the step encompassing identification of variable genes

(FindVariableFeatures). Integration anchors were then identified using the FindIntegrationAnchors and used to integrate all human fetal datasets through the IntegrateData function. Following integration, data were scaled (ScaleData), significant PCs were identified (RunPCA), and UMAP-based clustering was performed (RunUMAP, FindNeighbors, FindClusters). Gene expression of specific genes were visualized by using read levels from the "RNA" slot of the integrated Seurat object (accessed by inputting "rna_gene" into gene parameter).

Pseudotemporal ordering

For the pseudotemporal ordering analysis of our 12wpc_1 sample, we utilized Monocle v2.99.3 (named Monocle 3 alpha). Variable genes from the Seurat3 analysis of our 12wpc_1 samples (resolution 0.8) were used as input into Monocle, utilizing the VGAM::negbinomial.size expressionFamily, and clusters were projected onto the minimum spanning tree after ordering. The beginning of pseudotime was assigned using the function orderCells based on *NGN3* expression.

To conduct alpha and beta branch analysis, clusters along each branch were isolated and loaded into Monocle separately. Genes that changed significantly as a function of pseudotime were identified with Monocle's differentialGenetest function, and those that displayed a q-value less than 0.001 were selected for downstream analysis. These genes were then plotted as a heatmap (using plot_pseudotime_heatmap) that clustered genes based on similarities in expression patterns along pseudotime. The expression of individual genes was plotted using Monocle's plot_genes_in_pseudotime function.

For the pseudotemporal ordering analysis of our merged human fetal and hESC-derived cell datasets, we utilized Monocle3 v0.1.0 (named Monocle 3 beta). We used this version of

Monocle3 given its internal batch correction capabilities. For our merged human fetal pseudotemporal ordering analysis, variable genes from the Seurat3 integration analysis were used as input into Monocle. Clusters were projected onto the minimum spanning tree after ordering.

For our merged hESC-derived analysis, variable genes from CHGA+ sub-clustering were used as input into Monocle. To batch correct based on sample type, the residual_model_formula_str was set to "~ orig.ident" during the pre-process_cds step. To conduct branch-specific analyses, the choose_cells function was used to manually select the branches of interest in Monocle's graphical user interface. Once branches were selected, genes that changed significantly along pseudotime were identified using the graph_test function. Genes of interest were plotted along the Monocle trajectory using the plot_cells function.

Genetic engineering of the FEV-KO hESC line

The HUES8 hESC line was used to generate our FEV-KO line. For the FEV-KO hESC line, the FEV-KO gRNA (5'-CTGATCAACATGTACCTGCC-3') was designed on Benchling software and ordered from Dharmacon in a lyophilized format. The gRNA was suspended nuclease-free 10mM Tris-HCl Buffer (pH 7.4) ordered from Dharmacon (cat. no. B-006000-100) and stored as aliquots in -80C. HUES8 hESCs were grown on Matrigel-coated tissue culture plates, and on the morning of nucleofection, media was changed to mTeSR1 + 10 μM Rock inhibitor Y-27632 for 2 hours prior to nucleofection. Following this incubation step, hESCs were lifted from Matrigel plates and dissociated into a single-cell suspension using TrypLE Express. Cells were incubated in TrypLE Express dissociation reagent for 6 mins at RT. mTeSR1 + 10 μM Rock inhibitor Y-27632 was used to neutralize the dissociation, and cell suspension was filtered through a 37um filter.

To carry out the nucleofection, we mixed 2.75uL of tracrRNA (160uM) and 2.75uL of our FEV-KO gRNA (160uM) (to make the "RNA-complex") in a PCR strip tube and incubated for 30 mins in the 37°C cell culture incubator. After 30 mins, 5.5uL of purified Cas9-NLS protein (QB3 UC Berkeley MacroLab) was added to the RNA complex, gently mixed to make the RNP (ribonucleoprotein), and incubated at 37°C for exactly 15 mins. After exactly 15 mins, previously dissociated cells were resuspended in Lonza's P3 buffer from the P3 Primary Cell 4D-Nucleofector X Kit S (V4XP-3032). 10uL of cell suspension containing 400K cells were pipeted into one well of the Lonza nucleofection strip, and 10uL of the RNP was added. The nucleofection strip was then inserted into the Lonza 4D-Nucelofector (Lonza, AAF-1002B) and nucleofected with the CA137 setting compatible with the P3 buffer. Nucleofected cells were then transferred to a 15mL conical with 3mL of mTeSR1 containing 10 μM Rock inhibitor Y-27632 and pen/strep. Cell viability was determined via Moxiflow, and cells were plated in one well of a 6-well plate coated with Matrigel. Cells were grown for 2-3 passages in mTeSR1 containing 10 μM Rock inhibitor Y-27632 and pen/strep to allow for recovery from nucleofection.

To determine genomic editing efficiency of the FEV-KO nucleofection experiment, genomic DNA from nucleofected cells was harvested in QuickExtract DNA Extraction (Lucigen, QE09050) and then used for PCR amplification. The following forward and reverse primers targeting the FEV-KO editing site were used to produce a 491-bp amplicon: 5'-CCGTCTTCTCCTCCTTGTCACC-3' and 5'-CTCGGCCACAGAGTACTCCAC-3'. This amplicon is GC-rich, requiring use of a PCR polymerase capable of handling GC-rich amplicons (PrimeSTAR GXL Premix, Clontech). This DNA amplicon (and a wildtype DNA amplicon) was sent to Quintarabio for Sanger sequencing. The chromatographs of each sequencing run were used for TIDE (Tracking of Indels by Decomposition) analysis, which estimates the frequency of insertions and deletions (indels) in a pool of cells that has undergone genomic editing (https://tide.deskgen.com) (Brinkman et al., 2014). Cutting efficiency of hESCs nucleofected with FEV-KO gRNA was then determined.

To derive a clonal FEV-KO line from this heterogeneous pool of hESCs that have no mutation in the *FEV* locus, a mutation(s) in one *FEV* allele, or mutations on both *FEV* alleles, we clonally plated these cells on Matrigel-coated plates. Approximately 1.5K cells were dispersed onto a 10 cm Matrigel-coated plate and allowed to grow for 9-10 days in mTeSR1. For the first 4-5 days of culture, cells were cultured in mTeSR1 containing 10 μM Rock inhibitor Y-27632. Clonal colonies were then hand-picked under a colony-picking microscope under sterile conditions. These hand-picked colonies were each transferred into one well of a 96-well plate, allowed to grow for 2-3 days, and then successively passaged into large plate formats (96-well to 24-well to 6-well to 10cm dish). Clonality was first determined through TIDE analysis as outlined above and confirmed with TOPO cloning of the FEV-KO PCR amplicon.

Genetic engineering of the FEV-KI hESC lines

The HUES8 hESC line was used to generate our FEV-MYC, FEV-GFP, and FEV-tNGFR lines. The MYC, GFP, and tNGFR inserts were all commercially synthesized as gene blocks from Integrated DNA Technologies. 5' and 3' FEV locus homology arms that were 400bp in length were then added to each of the MYC, GFP, and tNGFR gene blocks using In-Fusion HD Cloning (Clontech, 638920). These homology arms flanked the cut site targeted by our FEV-KI gRNA. The result of In-Fusion HD cloning was a pUC19 plasmid containing a MYC, GFP, or tNGFR insert flanked by 5' and 3' *FEV* homology arms. These plasmids were transformed into Stellar Competent Cell (Clontech, 636766), and PCR amplification off of these isolated plasmids generated a PCR amplicon for use as our targeting template to knock in MYC, GFP, and tNGFR into the *FEV* locus. The following forward and reverse primers were used in PCR to generate each targeting template from each plasmid: 5'-TGAACTACGACAAGCTGAGCCG-3' and 5'-TCCTTGGGGAAGAGCAAAAGTG-3'.

For knock-in of MYC, GFP, and tNGFR into the FEV locus, a FEV-KI gRNA (GCCATTACCACTAGACGGGG) was designed using Benchling software and targeted the end of exon 3 of the FEV locus. This FEV-KI gRNA cut immediately preceding the *FEV* stop codon found at the end of exon 3 and would facilitate the knock-in of each insert in-frame with the *FEV* locus. On the morning of nucleofection, HUES8 hESCs were fed with mTeSR1 + 10 μM Rock inhibitor Y-27632 for 2 hours. Following this incubation step, hESCs were lifted from Matrigel-coated plates and dissociated into a single-cell suspension using TrypLE Express. Cells were incubated in TrypLE Express dissociation reagent for 6 mins at RT. mTeSR1 + 10 μM Rock inhibitor Y-27632 was used to neutralize the dissociation, and cell suspension was filtered through a 37um filter.

To carry out the nucleofection, we mixed 1.25uL of tracrRNA (160uM), 1.25uL of our FEV-KI gRNA (160uM), and 1ug of either the MYC, GFP, or tNGFR targeting templates in a PCR strip tube and incubated for 30 minutes in the 37C cell culture incubator. After 30 mins, 2.5uL of purified Cas9-NLS protein (QB3 UC Berkeley MacroLab) was added, gently mixed to make the RNP complex, and incubated at 37C for exactly 15 minutes. Dissociated cells were pelleted at 1000rpm for 3 mins and resuspended in Lonza P3 buffer (Lonza, V4XP-3032). 10uL of cell suspension containing 400K cells were then pipeted into a Lonza cuvette, and 10uL of the RNP complex + targeting template was added. The cuvette was then inserted into the Lonza 4D-Nucelofector (Lonza, AAF-1002B) and nucleofected with the CA137 setting compatible with the P3 buffer. Nucleofected cells were then transferred to a 15mL conical with 3mL of mTESR containing 10 μM Rock inhibitor Y-27632 and pen/strep. Cell viability was determined via Moxiflow, and cells were plated in one well of a 6-well plate coated in Matrigel. Cells were grown for 2 passages to allow for recovery from nucleofection.

To determine if the MYC, GFP, and tNGFR inserts were successfully knocked-in, genomic DNA from nucleofected cells was harvested in QuickExtract DNA Extraction (Lucigen, QE09050) and used for PCR amplification. The following forward and reverse primers were used:
MYC: 5'-AGATCCAGCTGTGGCAGTTTCT-3' and 5'-ACCAGACAAGGATTGAGGGAGC-3'
GFP: 5'-CGTGCATCTGGAAAGCTACGTG-3' and 5'-CTTGAAGAAGTCGTGGCGCTTC-3'
tNGFR: 5'-TGAACTACGACAAGCTGAGCCG-3' and 5'-TCCTTGGGGAAGAGCAAAAGTG-3'
Presence of a knock-in band that was larger than the *FEV* wildtype band was indicative that a subset of nucleofected cells carried the insert.

To derive clonal FEV-KI lines from a heterogeneous pool of hESCs that either had the knock-in insert or not, we clonally plated these cells on Matrigel-coated plates. Approximately 1.5K cells were dispersed onto a 10cm Matrigel plate and allowed to grow for 9-10 days in mTeSR1. For the first 4-5 days of culture, cells were cultured in mTeSR1 containing 10 μM Rock inhibitor Y-27632. These hand-picked colonies were each transferred into one well of a 96-well plate, allowed to grow for 2-3 days, and then successively passaged into large plate formats (96-well to 24-well to 6-well to 10cm dish). Genomic DNA was isolated from each clonal line and the insert was confirmed through PCR using same primers as indicated above. Sanger sequencing of the genomic FEV locus confirmed that the MYC, GFP, and tNGFR had no mutations and was in-frame with the endogenous *FEV* locus.

Generation of gene KOs during directed differentiation of hESCs

Approximately 100-150e6 End Stage 4 (ES4) cells from our directed differentiation protocol were dissociated in Accumax for 15-25 mins in a 37°C water bath. The dissociated cell suspension was passed through a 37um filter. Cell count and viability were determined with a Moxiflow cell counter. Cells were pelleted at 1000rpm for 3 mins and kept in ES4 media until nucleofection.

For nucleofection, we used the large format of Lonza's nucleofection kits (P3 Primary Cell 4D-Nucleofector X Kit L, V4XP-3024), which accommodates nucleofection of up to $20e^6$ cells per nucleofection vessel. Four conditions were typically included in these experiments: non-nucleofected control, scramble control, hAAVS1 control, and KO of gene of interest. The non-nucleofected control group contained ES4 cells that did not go through nucleofection. The scramble control group contained ES4 cells that were nucleofected with a scramble gRNA (GGTTCTTGACTACCGTAATT) that is not predicted to cut anywhere in the human genome. The hAAVS1 control included ES4 cells that were nucleofected with a gRNA targeting a safe harbor locus in the human genome AAVS1 (GGGGCCACTAGGGACAGGAT). The KO of gene of interest group contained ES4 cells that were nucleofected with a gRNA targeting the gene of interest we wished to knock out. All gRNAs were ordered from Dharmacon.

For each nucleofection set of 10-20e6 ES4 cells, we mixed 9.5uL of tracrRNA (160uM), 9.5uL of our FEV-KI gRNA (160uM) in a PCR tube and incubated for 30 minutes in the 37°C cell culture incubator. After 30 mins, 19uL of purified Cas9-NLS protein (QB3 UC Berkeley MacroLab) was added, gently mixed to make the RNP complex, and incubated at 37°C for exactly 15 minutes. Dissociated ES4 cells were pelleted at 1000rpm for 3 mins, and each set of 10-$20e^6$ ES4 cells were resuspended in 64uL of Lonza P3 buffer (from V4XP-3024). Each set of cells were then pipetted into a large Lonza nucleofection vessel, and 36uL of the RNP were added. Nucleofection vessel was then inserted into the Lonza 4D-Nucelofector (Lonza, AAF-1002B) and nucleofected with the CA137 setting compatible with the P3 buffer. Nucleofected cells were then transferred to a 15mL conical with 10mL of S5D1 media. Cell viability was determined via Moxiflow.

Following nucleofection, cells were immediately re-aggregated into clusters using AggreWell 400 plates (STEMCELL Technologies, 34415). Wells in the AggreWell 400 plates were washed with an Anti-Adherence Rinsing Solutions (STEMCELL Technologies, 07010) and centrifuged in a swinging bucket rotor at 1300 x g for 5 mins. Rinsing solution was removed, and S5D1 media was used to rinse wells. S5D1 media was aspirated, and $1.2e^6$ cells were then pipeted into each well of an AggreWell 400 plate. Plates were spun at 100 x g for 3 mins to facilitate re-aggregation of cells in each microwell and then were observed under microscope to verify even distribution of cells among microwells. Plates were placed in the 37°C cell culture incubator, and spheroids formed by 48 hours (by S5D3). On S5D3, clusters were removed from the AggreWell plates and cultured in either miniature spinner flasks called Biotts (BWV-S03A) set at a 70rpm rotation speed or in 6-well ultra low-attachment plates (5mL of media with approximately $5e^6$ cells per well) placed on an orbital shaker set to 100rpm. Directed differentiation of these nucleofected clusters was continued either in Biotts or in a 6-well ultra low-attachment plate.

FACS of hESC-derived cells

BLC clusters were dissociated in Accumax for 15-25 mins in a 37°C water bath. The dissociated cell suspension was passed through a 37um filter. Cells were pelleted at 1000rpm for 3 mins and fixed in 4% PFA for 12 mins at RT. Cells were washed in 1X PBS, pelleted again, and resuspended in 1X PBS. Fixed cells were stored in 4°C prior to staining for FACS.

For FACS staining, cells were permeabilized using 1X Permeabilization Buffer (Invitrogen, 00-8333-56) for 5 mins at RT. Cells were then incubated in primary antibody diluted in Blocking reagent (0.2% Triton X-100, 5% NDS, 1% Bovine Serum Albumin (BSA) in 1X PBS) overnight at 4C. Primary antibodies used were anti-Chromogranin A (1:500, Abcam ab15160) and anti-C-Peptide (1:200, EMD Millipore 05-1109). The next day, cells were washed in 1X Permeabilization Buffer for 5 mins at RT and incubated in species-specific Alexa Fluor 488- and

555-conjugated secondary antibodies (1:500, Jackson ImmunoResearch) for 30 mins at RT. Cells were then washed in 1X Permeabilization Buffer, pelleted, resuspended in 1X PBS, and analyzed with BD Fortessa Analyzer.

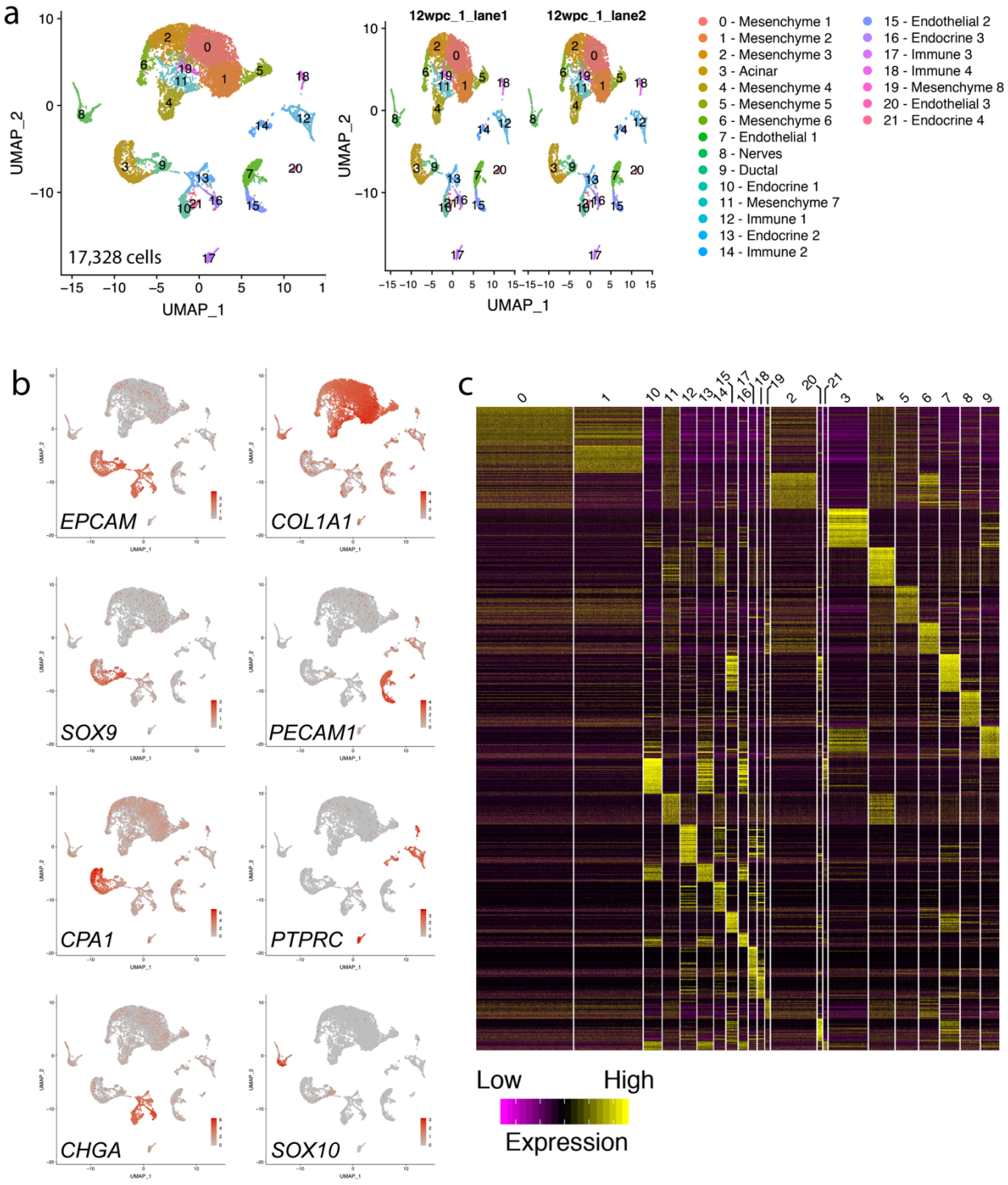


Figure 3.1. Single-cell RNA-sequencing identifies diverse cellular compartments in 12wpc human fetal pancreas. (a) UMAP-based clustering of single cells organized into 22 distinct clusters from one 12wpc human fetal pancreas. Each dot represents a single cell and is colored based on its assigned cluster identity. Plot to the left represents UMAP-based clustering of cells from each of two technical replicate samples run on two different wells of the 10X Chromium single-cell sequencing chip. Each technical replicate shows even contribution of cells to each

cluster within the merged UMAP. (b) Expression patterns of known marker genes of epithelial cells (CHD1+), ductal cells (SOX9+), acinar cells (CPA1+), endocrine cells (CHGA+), mesenchymal cells (COLA1A+), endothelial cells (PECAM1+), immune cells (PTPRC+), and nerve (SOX10+) cells, revealing the identities of all 22 clusters from UMAP-based clustering. Intensity of color indicates level of gene expression. (c) Heatmap showing the top 50 differentially-expressed genes in each cluster compared to all other clusters. Individual cells are represented in each column, and columns of cells derived from the same cluster are grouped together. Cluster numbers are consistent with cluster numbering system as shown in (a). Genes are represented by rows.

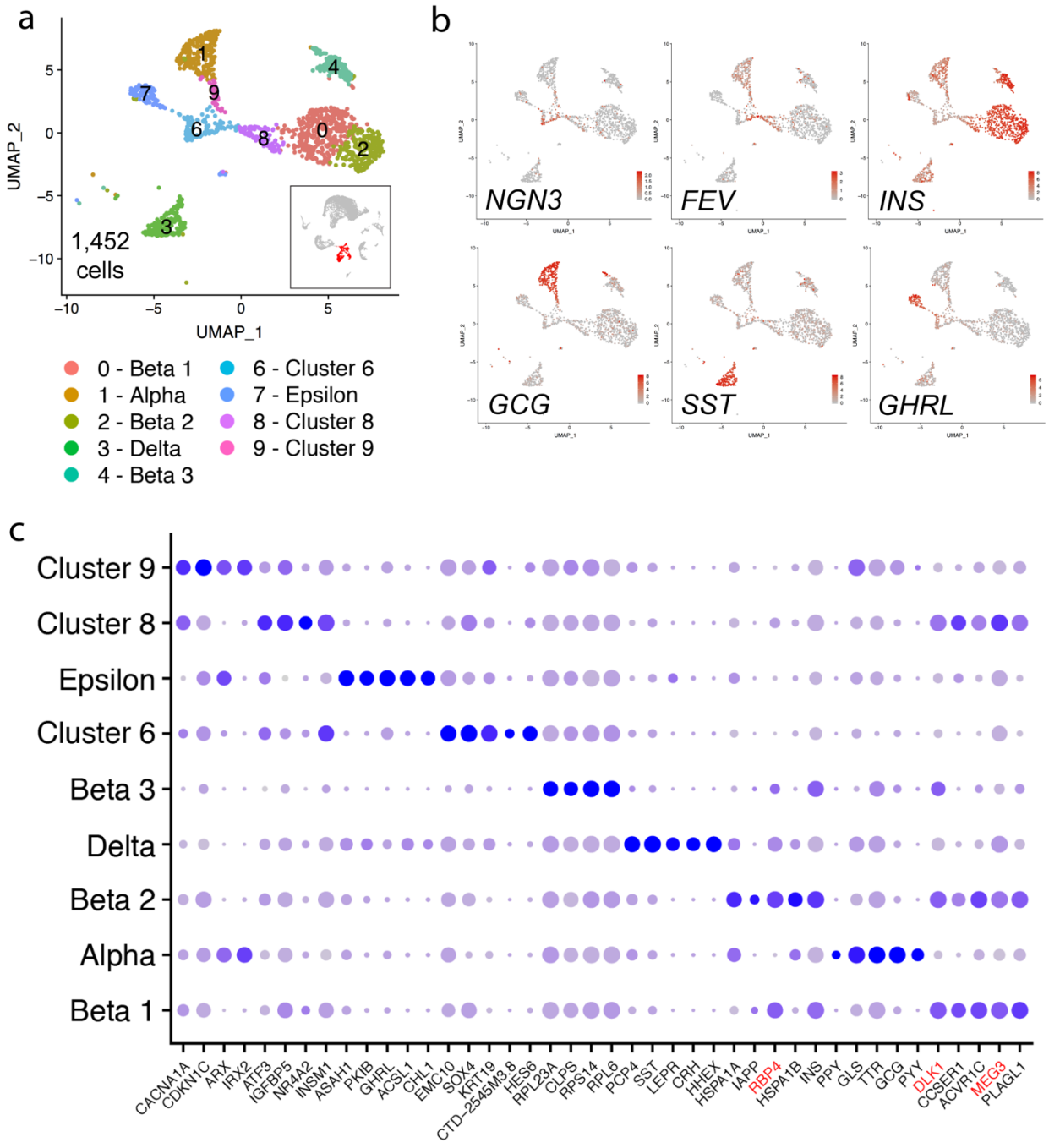


Figure 3.2. Endocrine sub-clustering identifies known and novel cell populations in 12wpc human fetal pancreas. (a) UMAP-based sub-clustering of CHGA+ clusters, as defined in Figure 17b, organized into 10 distinct populations. Inset shows CHGA+ clusters from Fig. 3.1b. (b) Known markers of the endocrine lineage, including *NGN3*, *INS, GCG, SST,* and *GHRL*, identifies NGN3+ endocrine progenitors, INS+ beta cells, GCG+ alpha cells, SST+ delta cells, and GHRL+ epsilon cells, respectively. *FEV* expression is also plotted to highlight novel populations previously uncharacterized in human pancreatic development. (c) Dot plot displaying the top five differentially-expressed genes from each cluster and their expression

levels across all 10 endocrine lineage clusters. Size of each dot represents the proportion of each population that expresses each specified gene. Color intensity reflects average level of gene expression. Genes highlighted in red are referred to in the body of the results section.

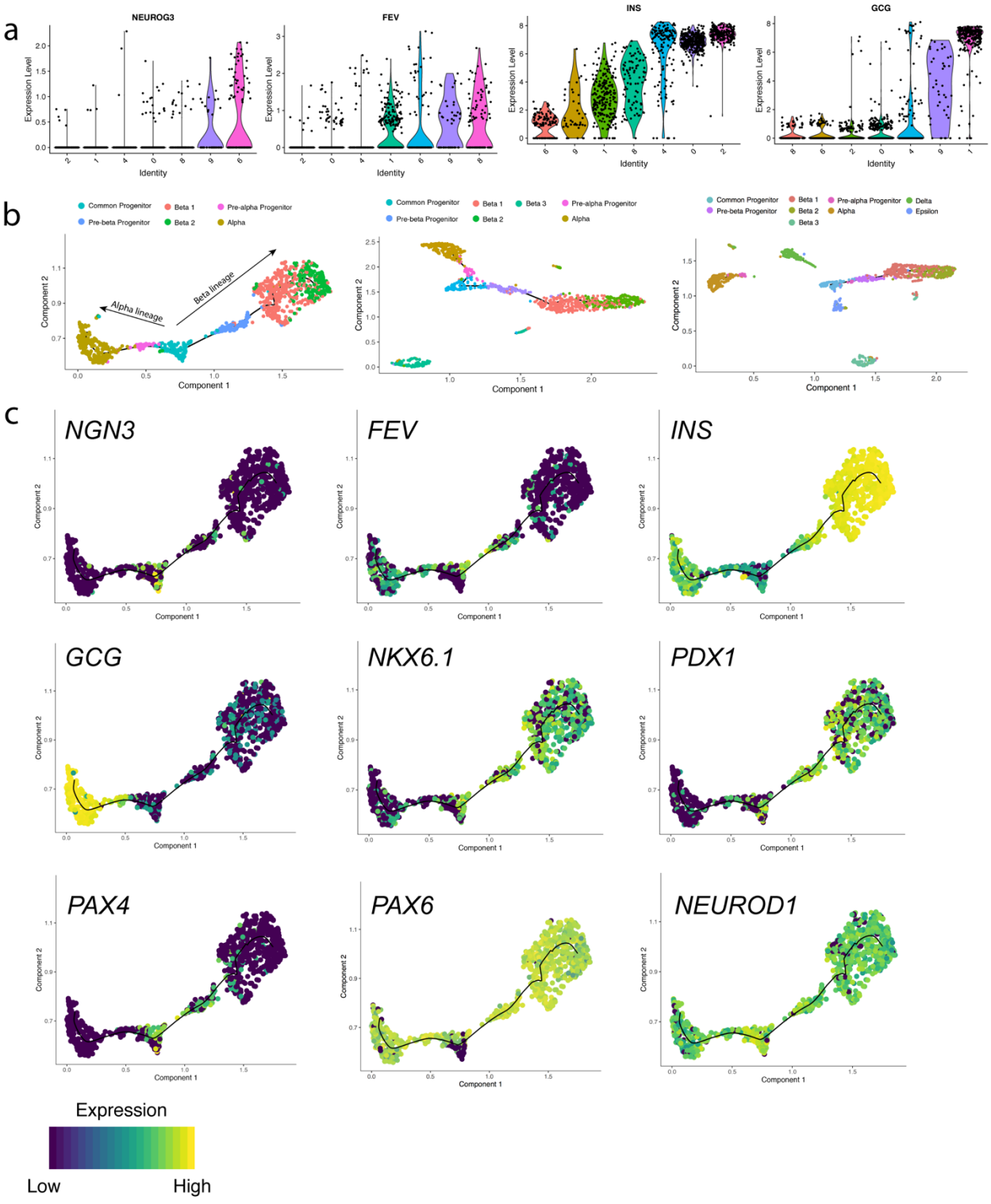


Figure 3.3. Identification of pre-beta and pre-alpha progenitors in 12wpc human fetal pancreas. (a) Violin plots depicting distribution of expression of *NGN3*, *FEV*, *INS,* and *GCG,* in single cells from clusters 6 (common endocrine progenitors), 8 (pre-beta progenitors), 0 (beta population #1), 2 (beta population #2), 4 (beta population #3), 9 (pre-alpha progenitors), and 1 (alpha population). Each dot represents the gene expression of a single cell, and the colored

distributions (“violins”) represent the spread of gene expression within each cluster. (b) Pseudotemporal ordering using Monocle 3 of common endocrine progenitors, pre-beta progenitors, pre-alpha progenitors, beta cells, and alpha cells (defined as clusters 6, 8, 9, 0, 2, and 1). Pseudotime begins at the vertex of the trajectory with common endocrine progenitors (cluster 6). Two differentiation arcs emanate from this cluster, leading to alpha and beta linages. Pre-beta progenitors (cluster 8) are placed immediately before differentiation into beta cells along pseudotime, and pre-alpha progenitors (cluster 9) are positioned immediately before differentiation into the alpha cell lineage. Second plot represents pseudotemporal ordering using the third beta cell population (Beta 3) as an additional input. Third plot represents pseudotemporal ordering of all endocrine lineages found in our 12wpc human fetal pancreas. (c) Gene expression intensity plots depicting gene expression in individual cells placed along pseudotime. Color intensity reflects level of gene expression.

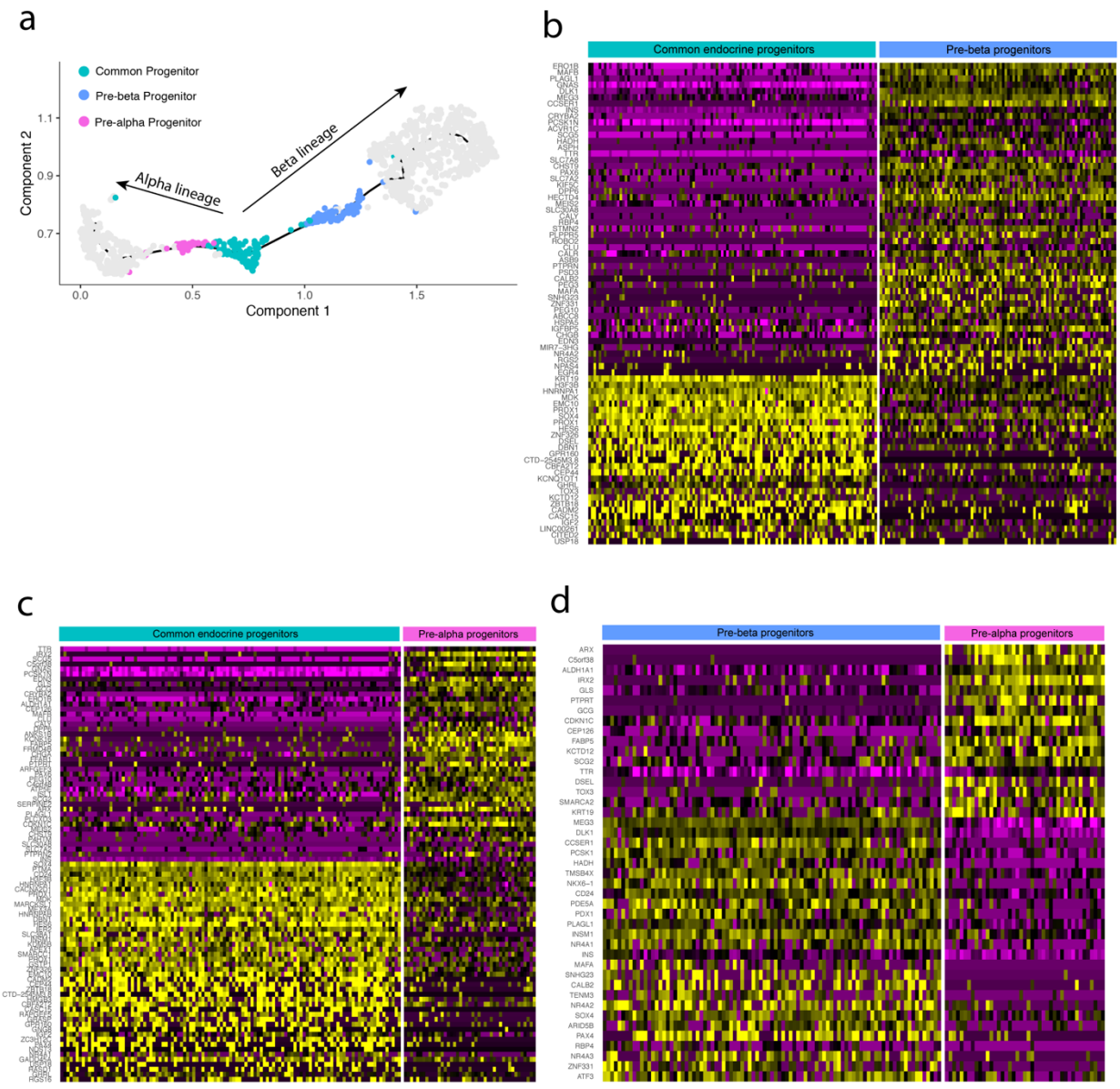


Figure 3.4. Transcriptomic profile comparison among pre-beta, pre-alpha, and common endocrine progenitors. (a) Pseudotemporal ordering highlighting cell populations (common endocrine progenitors, pre-beta progenitors, and pre-alpha progenitors) used for pairwise comparisons. (b-d) Pairwise comparisons among clusters 6 (common endocrine progenitors), 8 (pre-beta progenitors), and 9 (pre-alpha progenitors). Heatmaps depict the genes expressed 2-fold or greater in each comparison.

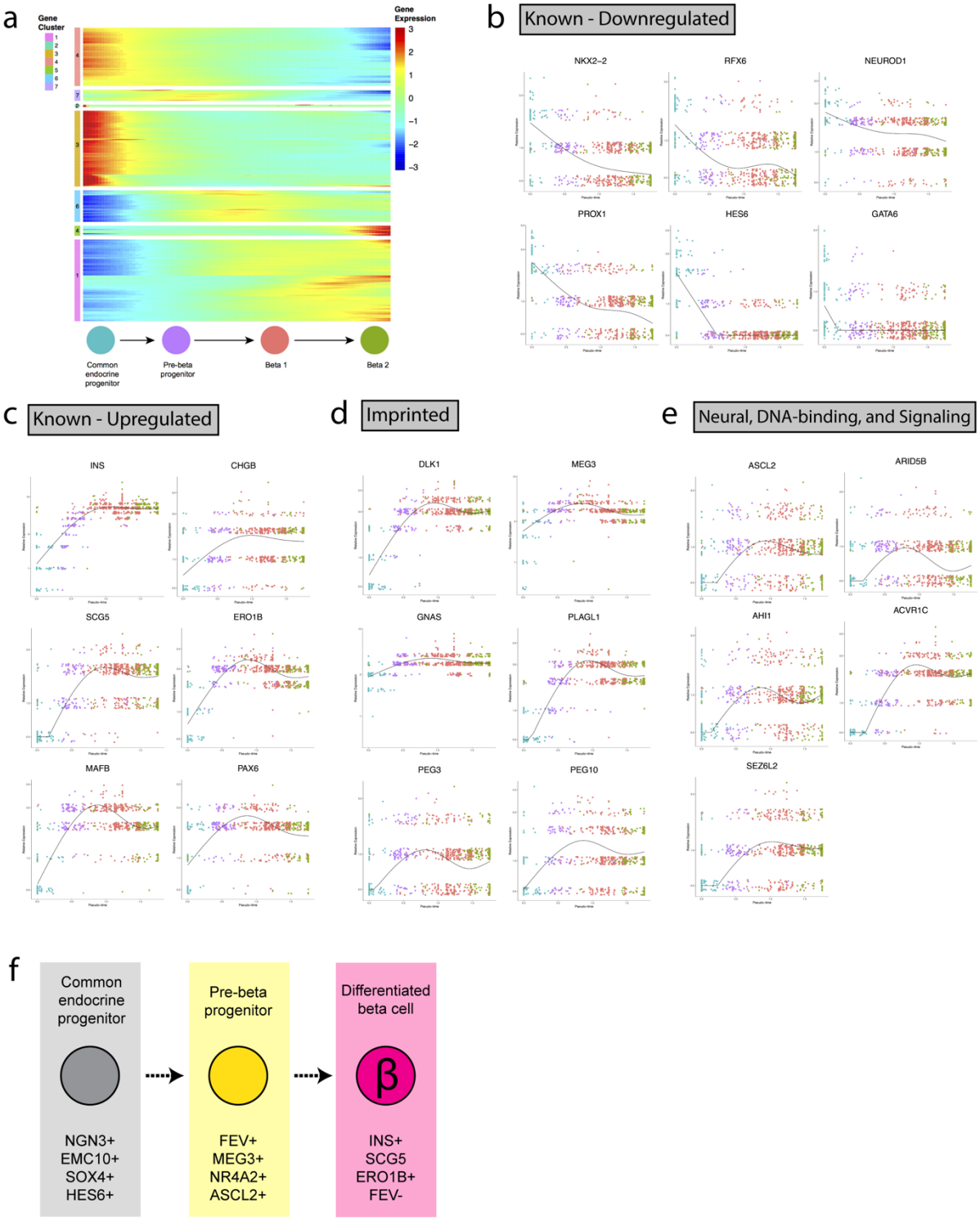


Figure 3.5. Identification of candidate regulators of beta lineage allocation in human endocrine cell development. (a) Heatmap depicting gene expression of cells (represented by columns) as a function of pseudotime across beta cell differentiation. Pseudotime begins with

common endocrine progenitors (cell cluster 6), followed by pre-beta progenitors (cell cluster 8), and ends with beta cell populations (cell clusters 0 and 2). Each individual row on the heatmap represents a gene, and the color intensity represents its expression along pseudotime. Genes that change significantly as a function of pseudotime are grouped in 7 main gene clusters: those that are highly expressed at the beginning of pseudotime (gene clusters 2, 3, and 4), genes that are upregulated during the pre-beta progenitor stage but taper in expression as beta cell identity is acquired (gene clusters 6 and 7), and genes that are upregulated during the pre-beta progenitor stage and remain expressed in the differentiated beta cell stage (gene clusters 1 and 5). (b-e) Pseudotime gene expression plots highlighting genes known to be downregulated as differentiation into the beta lineage occurs (b), genes known to be upregulated during beta cell differentiation (c), novel candidate regulators of beta cell lineage allocation that are imprinted (d), genes previously identified in the development and function of the nervous system (e), and those have known functions as a DNA-binding protein or in canonical signaling pathways, such as Activin A signaling (e). In (b-e), each dot represents the gene expression a single cell placed along pseudotime, and the color of the dot denotes its original cluster identity, which corresponds to the cell differentiation scheme outlined in (a). The black curve maps the average gene expression as a function of pseudotime. (f) Model for beta cell differentiation in human fetal pancreatic development.

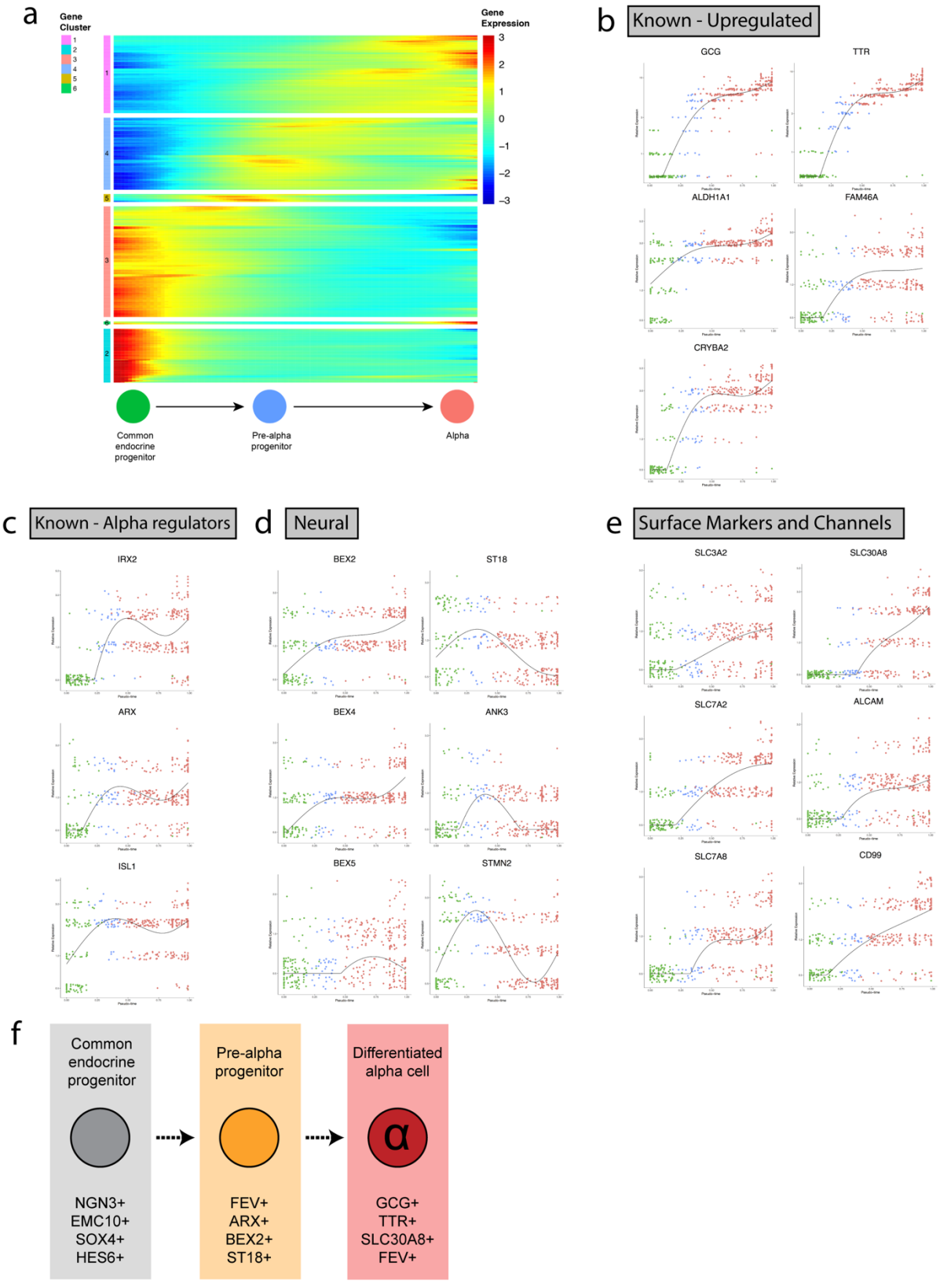

a
Gene Cluster
1
2
3
4
5
6
Gene Expression
3
2
1
0
−1
−2
−3
Common endocrine progenitor
Pre-alpha progenitor
Alpha
b
Known - Upregulated
GCG
TTR
ALDH1A1
FAM46A
CRYBA2
c
Known - Alpha regulators
IRX2
ARX
ISL1
d
Neural
BEX2
ST18
BEX4
ANK3
BEX5
STMN2
e
Surface Markers and Channels
SLC3A2
SLC30A8
SLC7A2
ALCAM
SLC7A8
CD99
f
Common endocrine progenitor
NGN3+
EMC10+
SOX4+
HES6+
Pre-alpha progenitor
FEV+
ARX+
BEX2+
ST18+
Differentiated alpha cell
α
GCG+
TTR+
SLC30A8+
FEV+

Figure 3.6. Identification of candidate regulators of alpha lineage allocation in human endocrine cell development. (a) Heatmap depicting average gene expression of cells as a function of pseudotime during alpha cell differentiation. Pseudotime begins with common endocrine progenitors (cluster 6), followed by pre-alpha progenitors (cluster 9), and ends with differentiated alpha cells (cluster 1). Each individual row on the heatmap represents a gene, and the color intensity represents its expression along pseudotime. Genes that change significantly as a function of pseudotime are grouped in 6 main gene clusters: those that are highly expressed at the beginning of pseudotime (gene clusters 2 and 3), genes that are upregulated during the pre-alpha progenitor stage but taper in expression as alpha cell identity is acquired (gene cluster 5), and genes that are upregulated during the pre-alpha progenitor stage and remain expressed in the differentiated alpha cell stage (gene clusters 1, 4, and 6). (b-e) Pseudotime gene expressions plots highlighting genes known to be expressed by the adult alpha cell lineage (b), genes known to be upregulated specifically during alpha cell differentiation (c), and novel candidate regulators of alpha cell lineage allocation that have previously been identified in the development and function of the nervous system (d) or are cell surface markers and channels (e). In (b-e), each dot represents the gene expression a single cell placed along pseudotime, and the color of the dot denotes its original cluster identity, which corresponds to the cell differentiation scheme outlined in (a). The black curve maps the average gene expression as a function of pseudotime. (f) Model for alpha cell differentiation in human fetal pancreatic development.

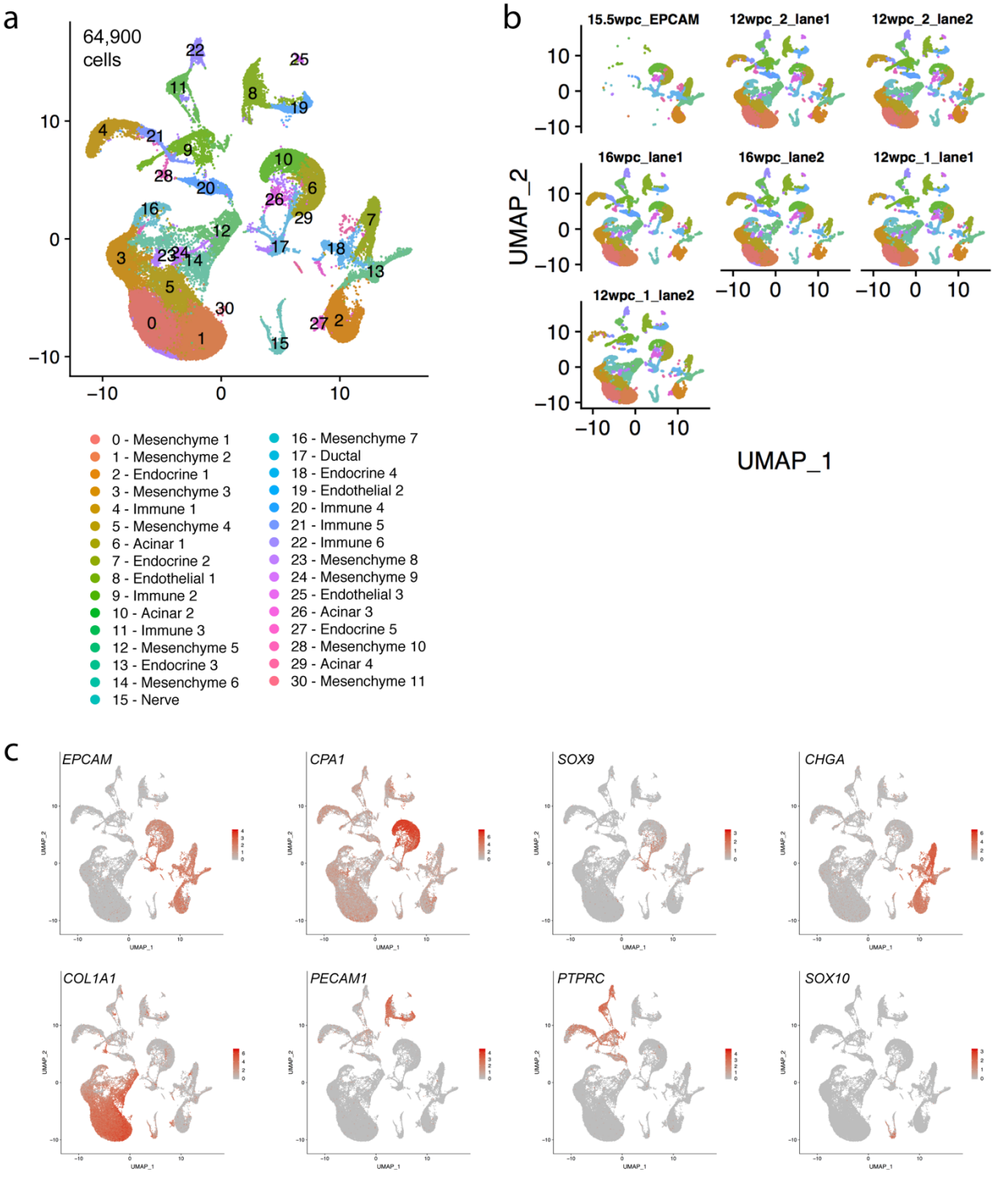


Figure 3.7. Human fetal pancreatic populations over developmental time. (a) UMAP-based clustering of merged cell populations from 12wpc_1, 12wpc_2, 15.5wpc, and 16wpc samples. Merging of these four datasets was accomplished using Seurat 3's Integration method. (b) UMAP clustering of each sample and technical replicate (if any) showing contribution of designated sample to overall merged UMAP clustering shown in (a). The 12wpc_1, 12wpc_2, and 16wpc samples were each run on two lanes of the 10X Chromium single-cell sequencing

chip, while the 15.5wpc sample, which was enriched for EPCAM+ cells, was run on only one lane of the 10X chip. (c) Known marker genes of epithelial cells (EPCAM+), ductal cells (SOX9+), acinar cells (CPA1+), endocrine cells (CHGA+), mesenchymal cells (COLA1A+), endothelial cells (PECAM1+), immune cells (PTPRC+), and nerves (SOX10+), showing the identities of all 31 clusters from UMAP-based clustering. Color indicates level of gene expression.

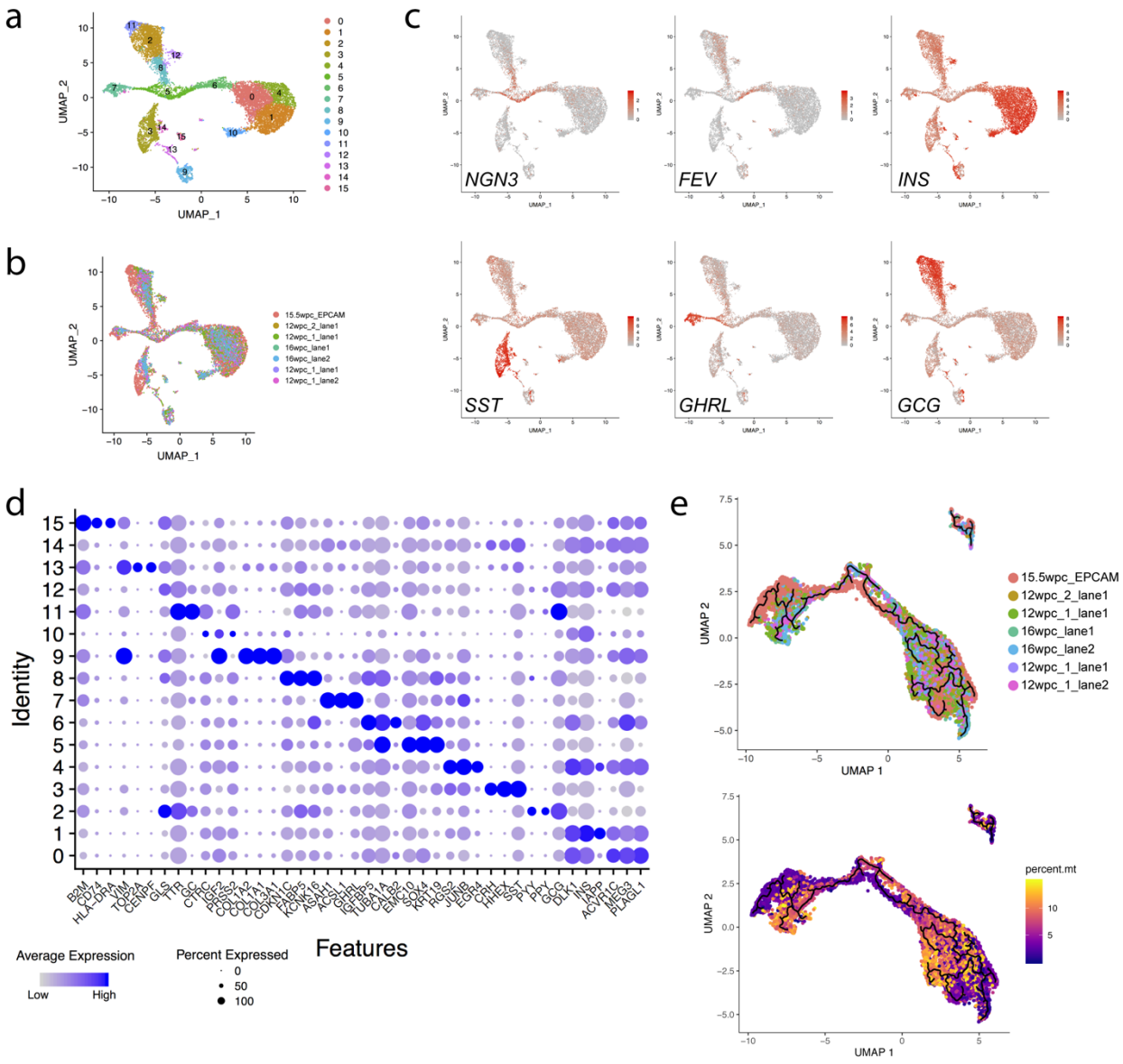


Figure 3.8. Human fetal endocrine populations over developmental time. (a) UMAP-based clustering resulting from endocrine sub-clustering of CHGA+ clusters from merged dataset (Fig. 3.7a) organized into 16 distinct populations. (b) Sample ID mapped onto UMAP depicts contribution of each timepoint (and associated technical replicate, if any) to each merged cluster. (c) Gene expression intensity plots highlighting NGN3+ progenitors, FEV+ cells, INS+ beta cells, GCG+ alpha cells, SST+ delta cells, and GHRL+ epsilon cells. (d) Dot plot displaying the top three differentially-expressed genes from each cluster and their expression levels across all 16 endocrine lineage clusters. Size of each dot represents the proportion of each population that expresses each specified gene. Color intensity reflects average level of gene expression. (e) Pseudotime ordering with Monocle 3 using endocrine progenitors and the alpha and beta lineages as input. Both the sample ID and mitochondrial content are mapped onto the pseudotime ordering trajectory, highlighting significant batch effect that is not corrected due to inclusion of 15.5wpc sample. 15.5wpc sample displays significantly lower mitochondrial content, which drives batch effect.

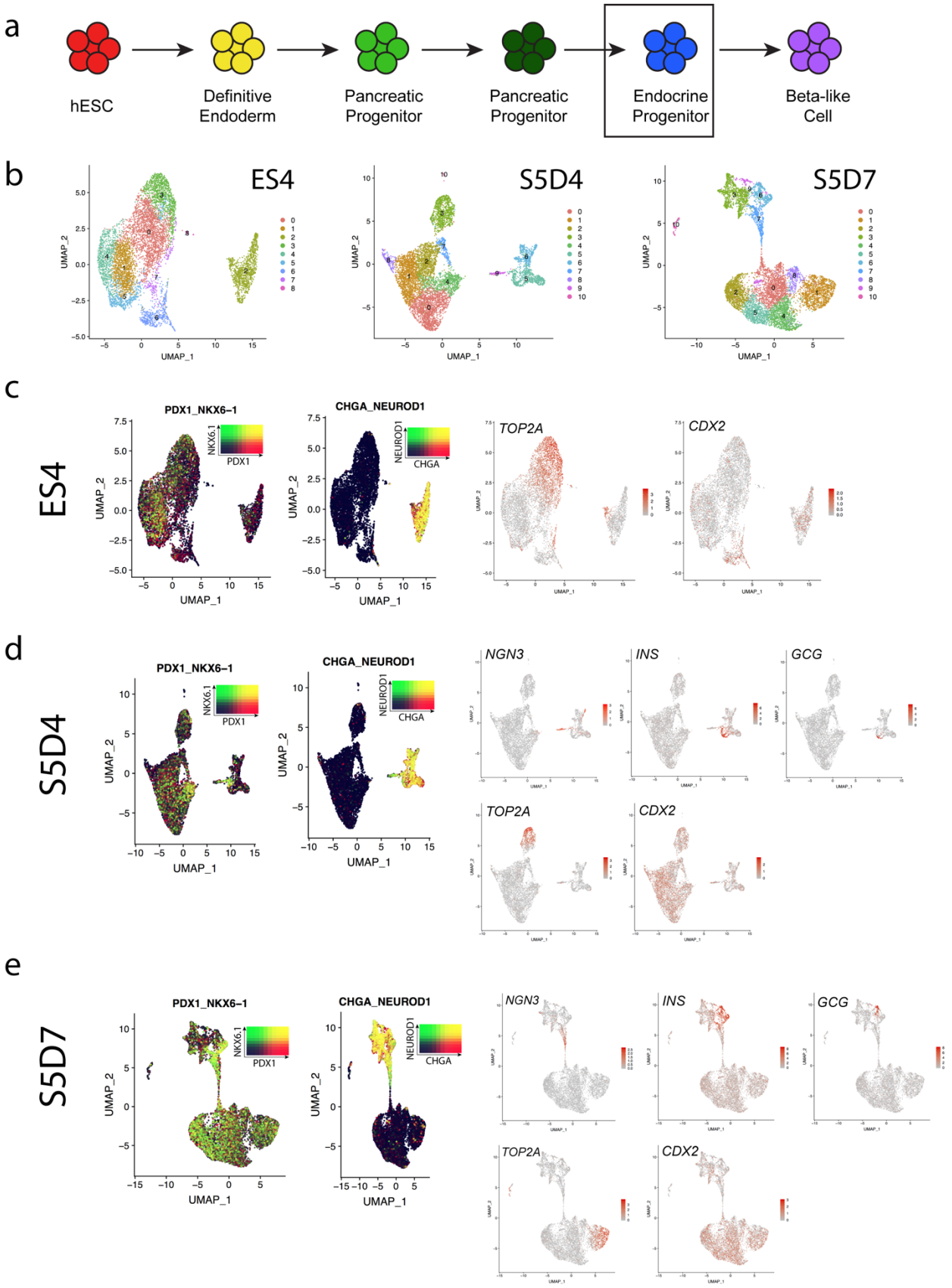

a
hESC
Definitive Endoderm
Pancreatic Progenitor
Pancreatic Progenitor
Endocrine Progenitor
Beta-like Cell
b
ES4
S5D4
S5D7
UMAP_1
UMAP_2
c
ES4
PDX1_NKX6–1
CHGA_NEUROD1
TOP2A
CDX2
NKX6.1
PDX1
NEUROD1
CHGA
d
S5D4
NGN3
INS
GCG
e
S5D7

Figure 3.9. Single-cell RNA-sequencing identifies heterogeneous cellular compartments in hESC-derived endocrine progenitor populations. (a) Schematic depicting the six stages of the protocol for *in vitro* beta cell differentiation, highlighting End Stage 4 pancreatic progenitors and Stage 5 endocrine progenitors taken for single-cell RNA-sequencing. (b) UMAP-based clustering of ES4 (end stage 4), S5D4, and S5D7 cells. (c-e) Blended expression plots highlight cells that express either both *PDX1* and *NKX6*-1 or *NEURDO1* and *CHGA*. Gene expression plots of *NGN3*, *TOP2A, CDX2, INS, and GCG* highlight endocrine progenitors, replicating cells, cells that mis-differentiated into a CDX2+ intestinal lineage, and hormone-expressing populations, respectively.

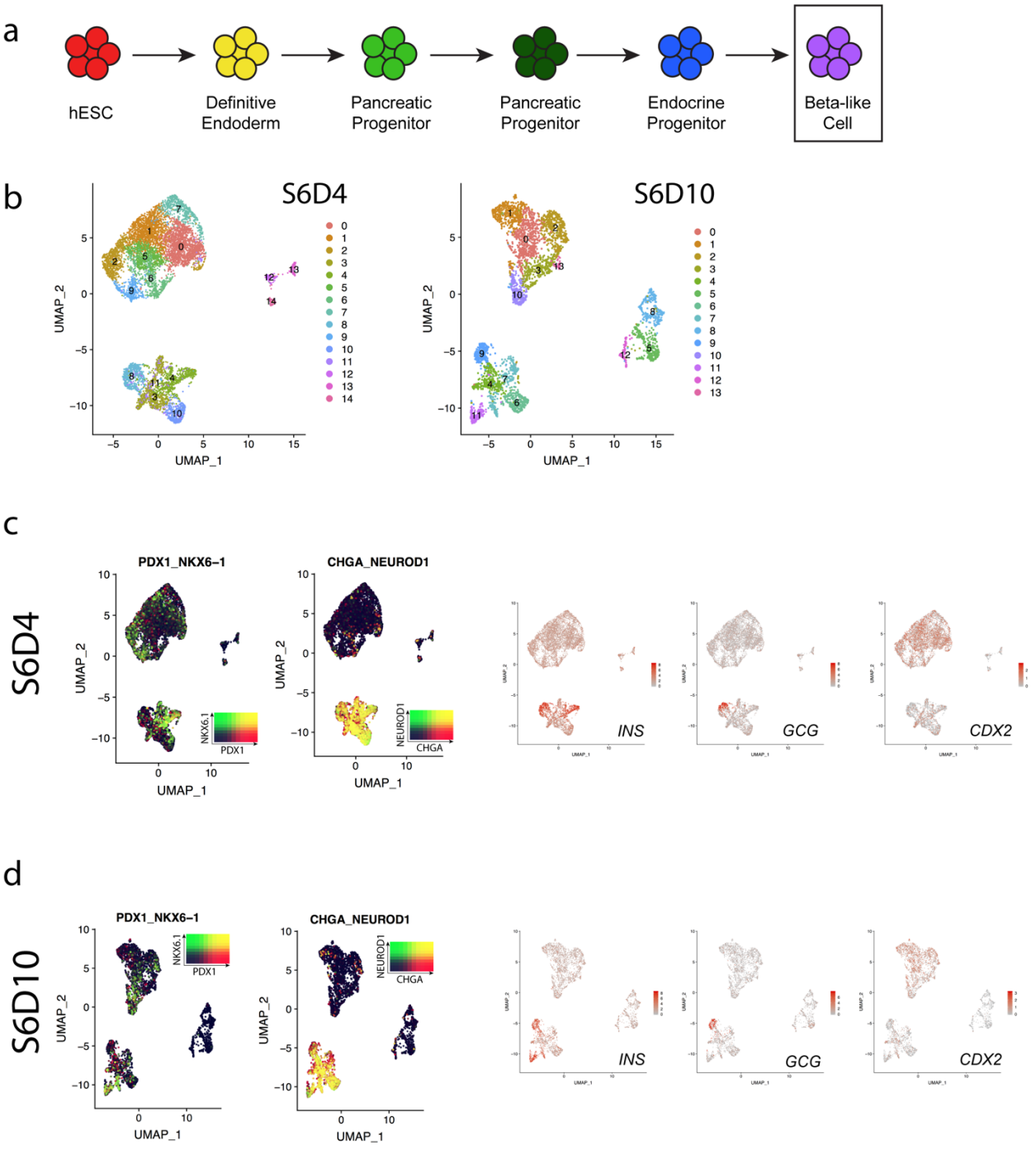


Figure 3.10. Single-cell RNA-sequencing identifies heterogeneous cellular compartments in hESC-derived beta-like stage cells. (a) Schematic depicting the six stages of our *in vitro* beta cell differentiation, highlighting Stage 6 beta-like cells taken for single-cell RNA-sequencing. (b) UMAP-based clustering of S6D4 and S6D10 beta-like stage cells organized into distinct clusters. (c, d) Blended expression plots highlight cells that express either both *PDX1* and *NKX6*-1 or *NEURDO1* and *CHGA*. Gene expression plots of *CDX2, INS, and GCG* cells that represent cells mis-differentiated into a CDX2+ intestinal lineage or hormone-expressing populations, respectively.

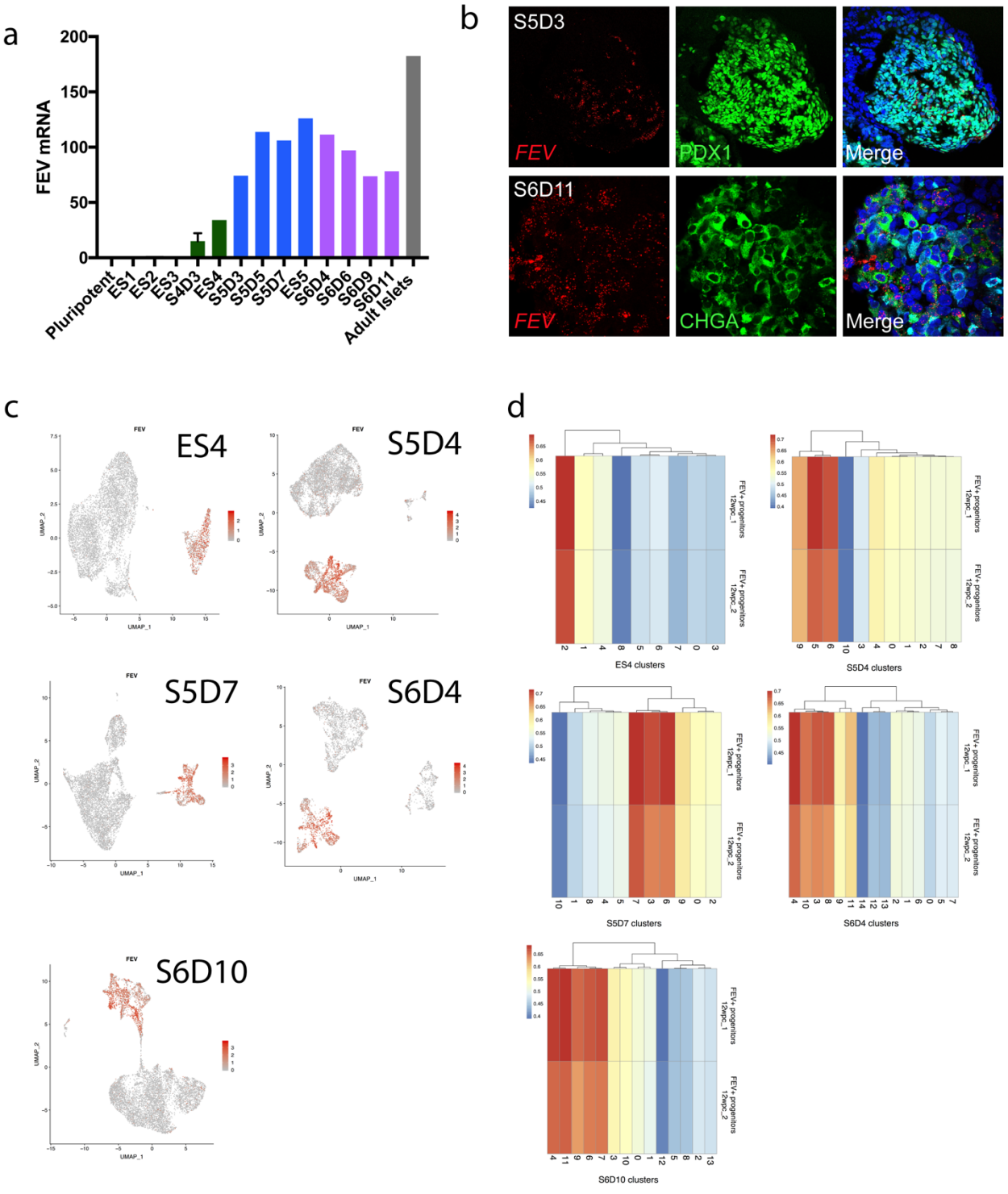


Figure 3.11. Emergence of FEV+ cells during *in vitro* beta cell differentiation. (a) qPCR (Taqman) data depicting *FEV* expression throughout the directed differentiation of hESCs towards the beta lineage. *FEV* expression in isolated adult human islets shown as a comparator. (b) Dual *in situ* hybridization and immunofluorescence on S5D3 and S6D11 hESC-derived clusters. *FEV* transcript is represented in red, and DAPI staining is in blue. For S5D3 clusters, green represents PDX1. For S6D11 clusters, green represents CHGA. (c) *FEV*

expression plots from ES4, S5D4, S5D7, S6D4, and S6D10 single-cell RNA-sequencing. (d) Heatmaps depicting the results of a Pearson's correlation analysis comparing FEV+ progenitors from 12wpc_1 and 12wpc_2 human fetal datasets with FEV+ clusters found in each sampled timepoint of directed differentiation. Color denotes level of correlation with FEV+ progenitors from 12wpc_1 and 12wpc_2 human fetal datasets, where shades of red reflect high transcriptional correlation.

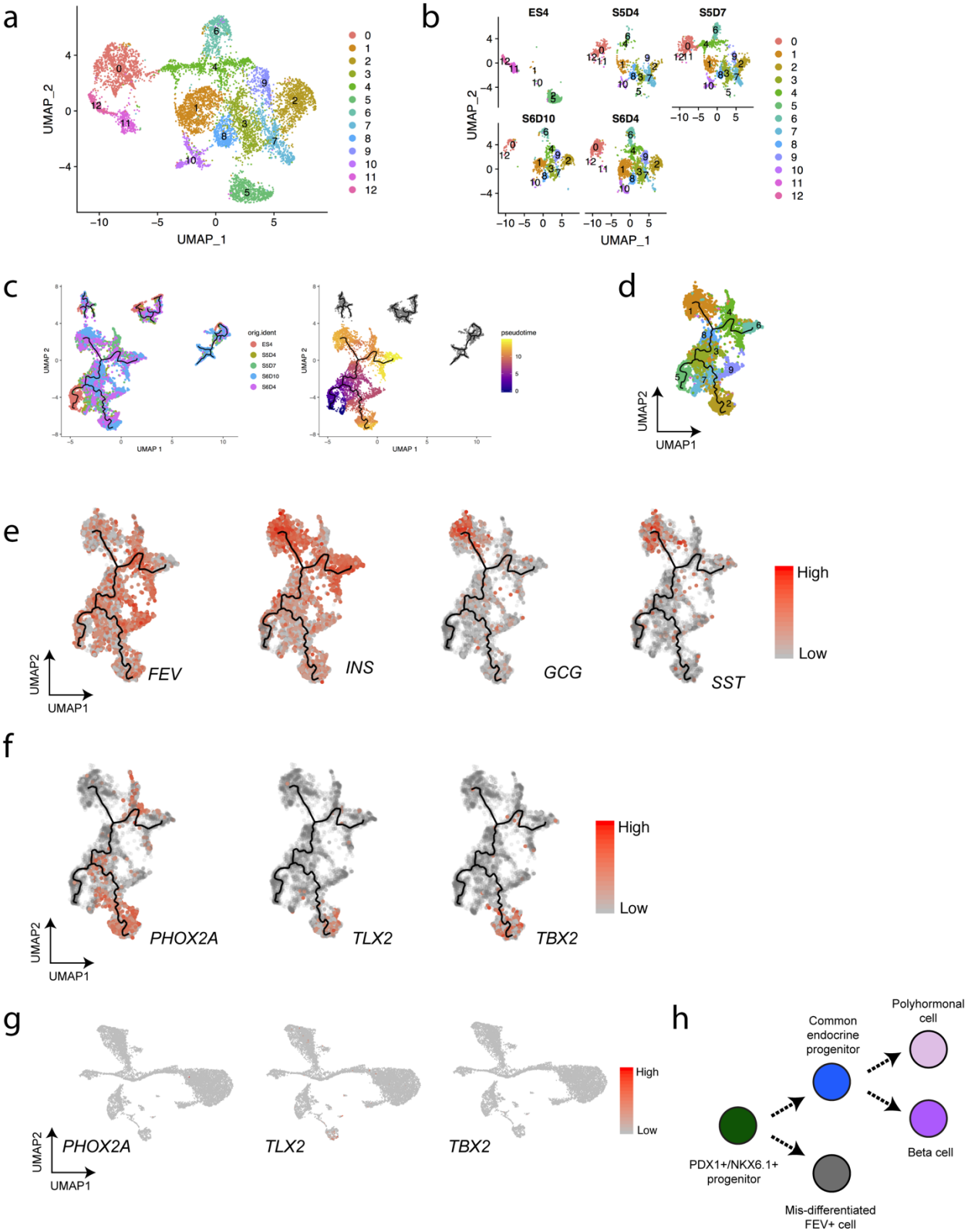


Figure 3.12. Reconstruction of lineage relationships among hESC-derived endocrine cells during *in vitro* beta cell differentiation. (a) UMAP-based clustering of merged CHGA+ clusters from ES4, S5D4, S5D7, S6D4, and S6D10 timepoints. (b) Timepoint IDs are mapped

onto UMAP clusters to illustrate the contribution of each timepoint to each resulting merged cluster. (c) Pseudotemporal ordering of the merged dataset using Monocle 3 depicts one main differentiation trajectory. Timepoint IDs mapped onto the trajectory reveals a cluster of ES4 cells at one end of the trajectory, which we designate as the beginning of our pseudotime ordering analysis. (d) Monocle trajectory with highlighted clusters labeled by their cluster identity as found in (a). (e) Gene expression intensity plots depicting *INS, GCG*, and *SST* expression in individual cells placed along pseudotime highlight the poly-hormonal and INS+ beta cell branches in the differentiation trajectory. *FEV* gene expression intensity plot depicts uniform *FEV* expression throughout the majority of pseudotime. (f) Gene expression intensity plots depicting *PHOX2A, TLX2,* and *TBX2* expression in individual cells placed along pseudotime highlight the restriction of these genes in the hormone-negative branch that contains the hESC-derived cells that are predicted to have mis-differentiated. (g) *PHOX2A, TLX2, and TBX2* expression plots from merged, endocrine sub-clustered UMAP from Fig. 3.8a. (h) Model of a bifurcation event of PDX1+/NKX6.1+ progenitors into a mis-differentiated FEV+ lineage and hormone-expressing endocrine lineages.

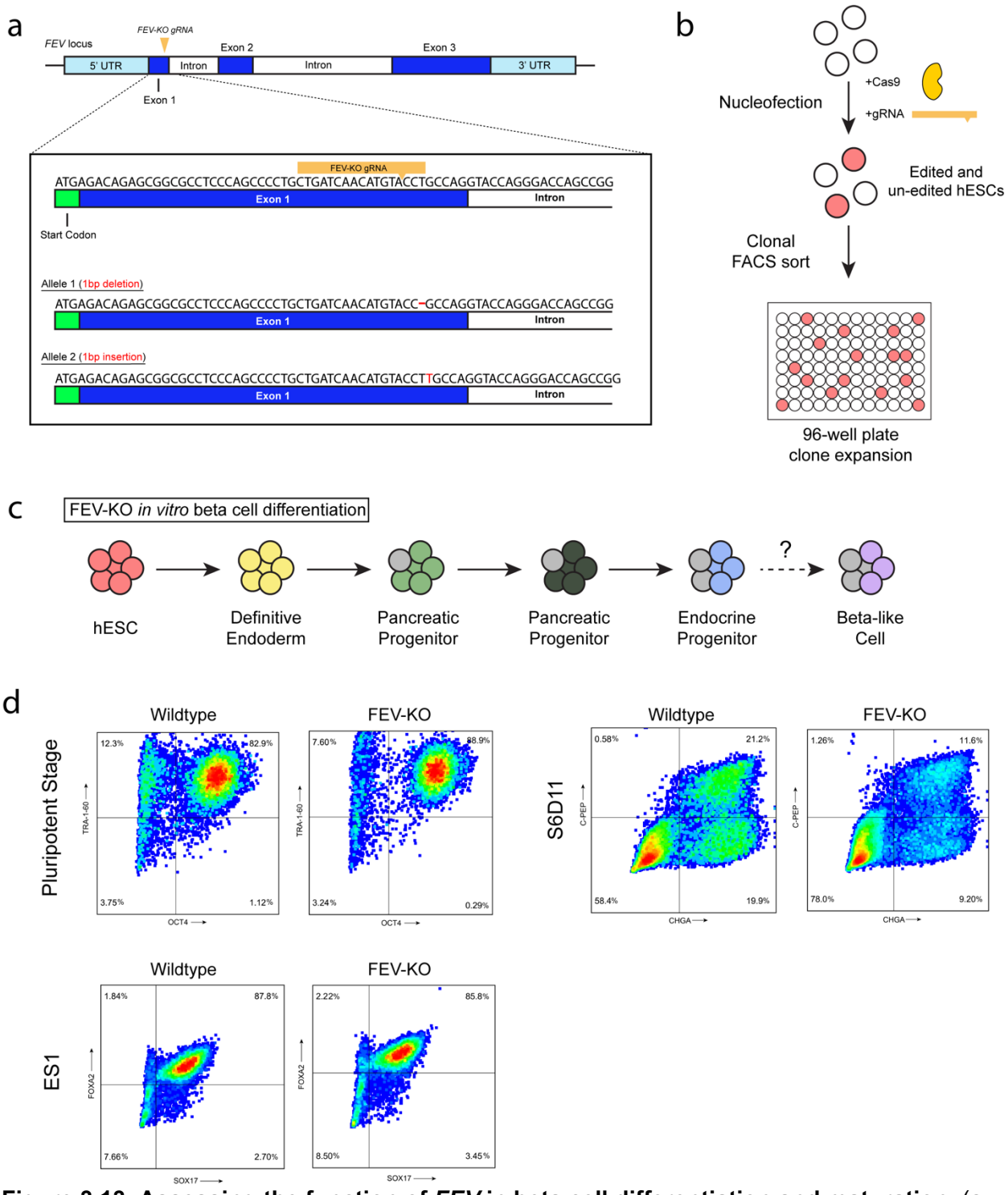


Figure 3.13. Assessing the function of *FEV* in beta cell differentiation and maturation. (a, b) Schematic illustrating the *FEV* locus and the use of CRISPR/Cas9-mediated genomic editing to generate a FEV-KO hESC clonal line. A FEV-KO gRNA was designed to target exon 1. Genomic editing with this FEV-KO gRNA led to the generation of a FEV-KO hESC clonal line with a 1-bp deletion in one allele and a 1-bp insertion in the second allele, leading a homozygous mutation in the *FEV* locus. (c) Illustration of the directed differentiation of FEV-KO hESCs towards a beta lineage. (d) FACS analysis of wildtype and *FEV-KO* pluripotent cells

quantifies percentage of cells with pluripotency markers OCT4 and TRA-1-60. FACS analysis of wildtype and *FEV-KO* cells at the completion of Stage 1 of the directed differentiation quantifies percentage of cells that have entered into the SOX17+/FOXA2+ definitive endoderm stage. FACS analysis of wildtype and *FEV-KO* cells at Stage 6, Day 11 (S6D11) determines the percentage of CHGA+/CPEP+ beta cells present.

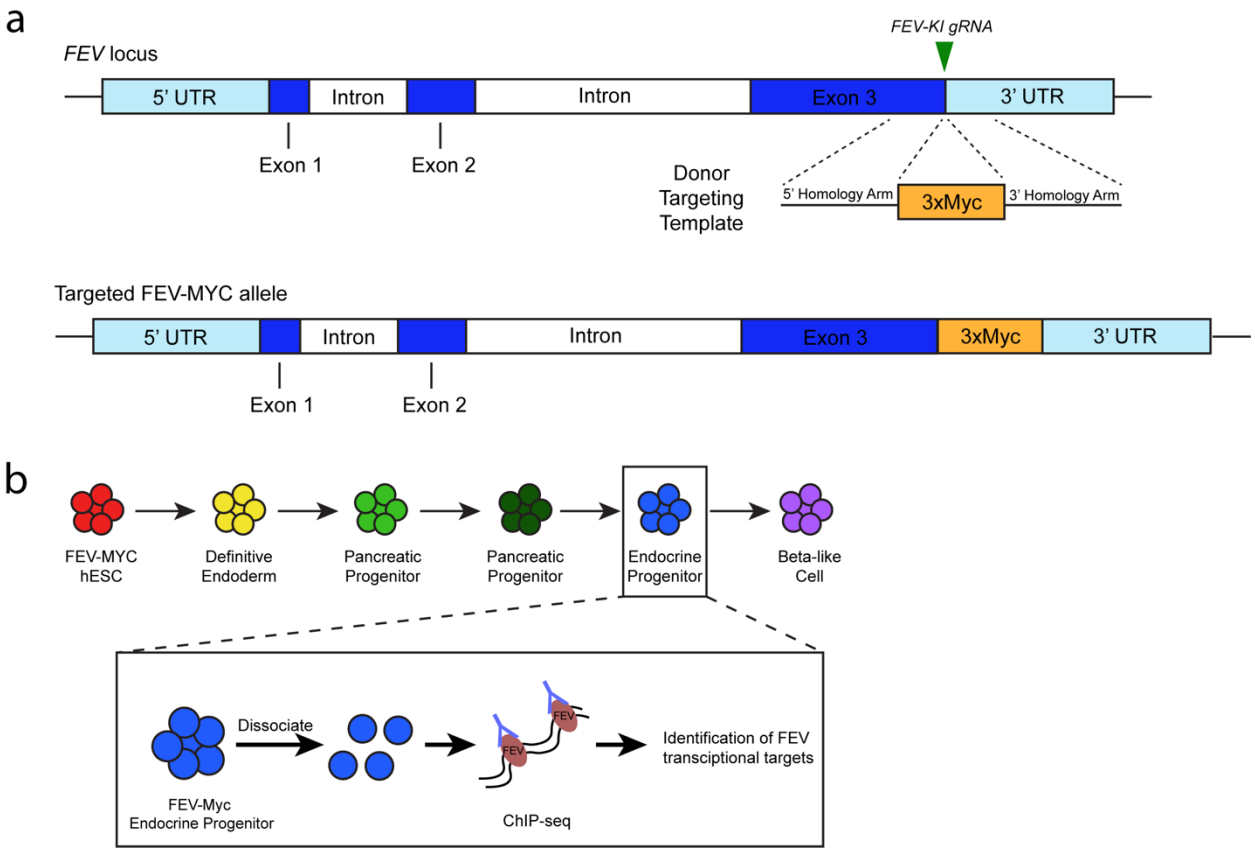


Figure 3.14. Strategy for identifying transcriptional targets of FEV. (a) Schematic illustrating the use of CRISPR/Cas9-mediated genomic editing to generate a FEV-MYC hESC line. A FEV-KI (knock-in) gRNA was designed to target the 3' end of exon 3 of the FEV locus. A targeting template containing a 3xMYC sequence flanked by homology arms was also designed and commercially synthesized. Through use of genomic editing and homology-directed repair, the 3xMYC sequence was knocked-in in frame with the endogenous FEV locus, leading to the expression of a FEV-MYC fusion protein to be used for ChIP-seq. (b) Illustration of future plans for the directed differentiation of FEV-MYC hESCs towards a beta lineage. Endocrine progenitor-stage cells will be harvested and ChIP-seq will be performed using an antibody against MYC. The MYC antibody will pull down FEV transcription factor bound to DNA, and sequencing of bound DNA will identify transcriptional targets of FEV specifically during this endocrine progenitor stage prior to beta cell lineage determination.

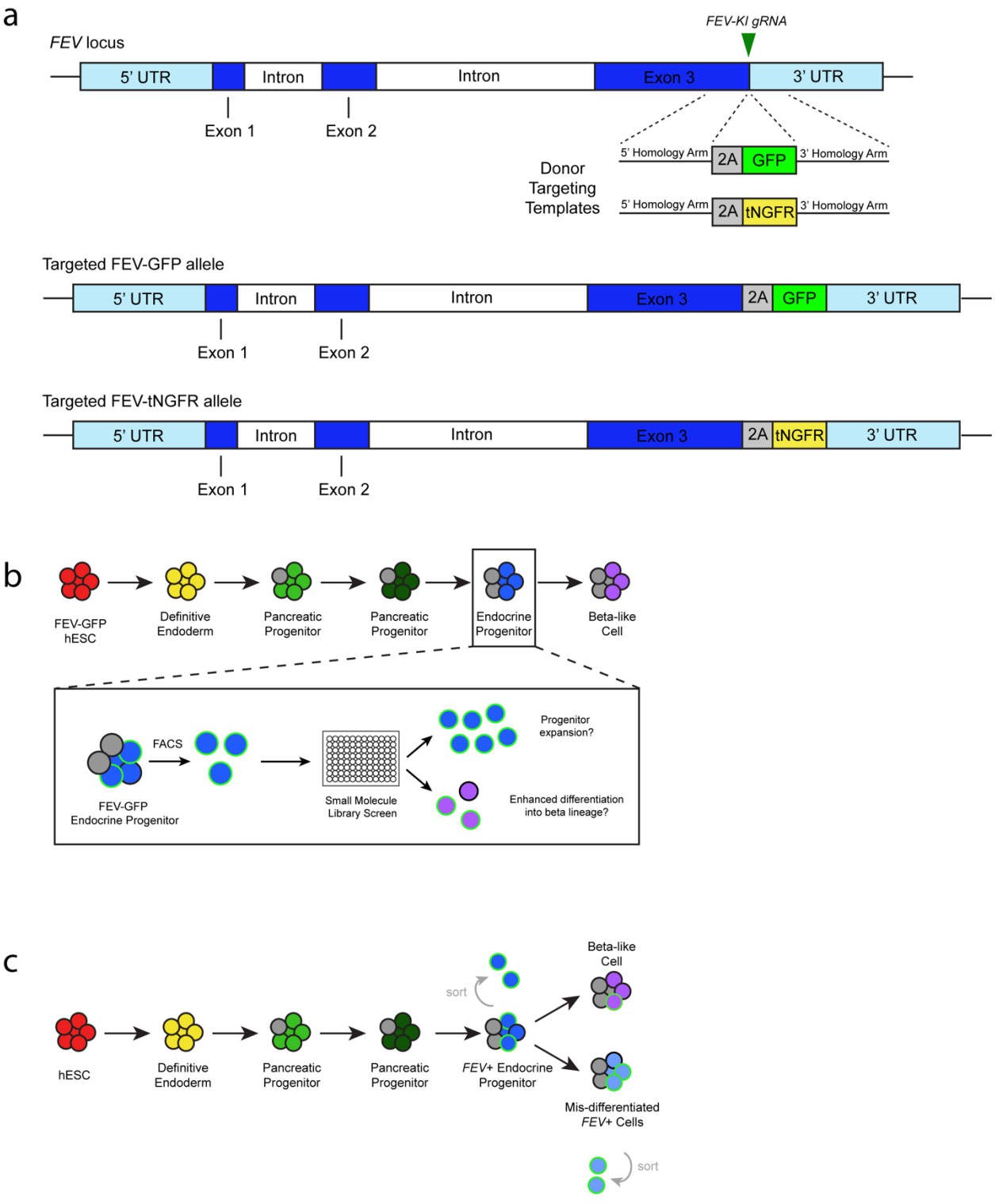


Figure 3.15. Identifying and isolating FEV-expressing cells during *in vitro* beta cell differentiation. (a) Schematic illustrating the use of CRISPR/Cas9-mediated genomic editing to generate two FEV reporter hESC lines: a FEV-GFP and a FEV-tNFGR line. A FEV-KI gRNA targets the 3' end of exon 3 of the *FEV* locus, and a targeting template containing either a T2A-GFP or T2A-tNGFR sequence flanked by homology arms was also designed and commercially

synthesized. Through use of genomic editing and homology-directed repair, the T2A-GFP or T2A-tNGFR sequence was knocked-in in frame with the endogenous FEV locus, leading to bicistronic translation of the FEV transcription factor and reporter protein (GFP or tNGFR). (b) Illustration depicting future use of our FEV reporter lines, such as our FEV-GFP line, to perform small molecule library screens to identify signaling compounds that induce FEV+ progenitor replication or enhance beta cell differentiation from FEV+ progenitors. (c) Illustration showing future use of our FEV reporter lines to sort and isolate different FEV+ populations, such as a predicted mis-differentiated FEV+ cell population, throughout *in vitro* beta cell differentiation.

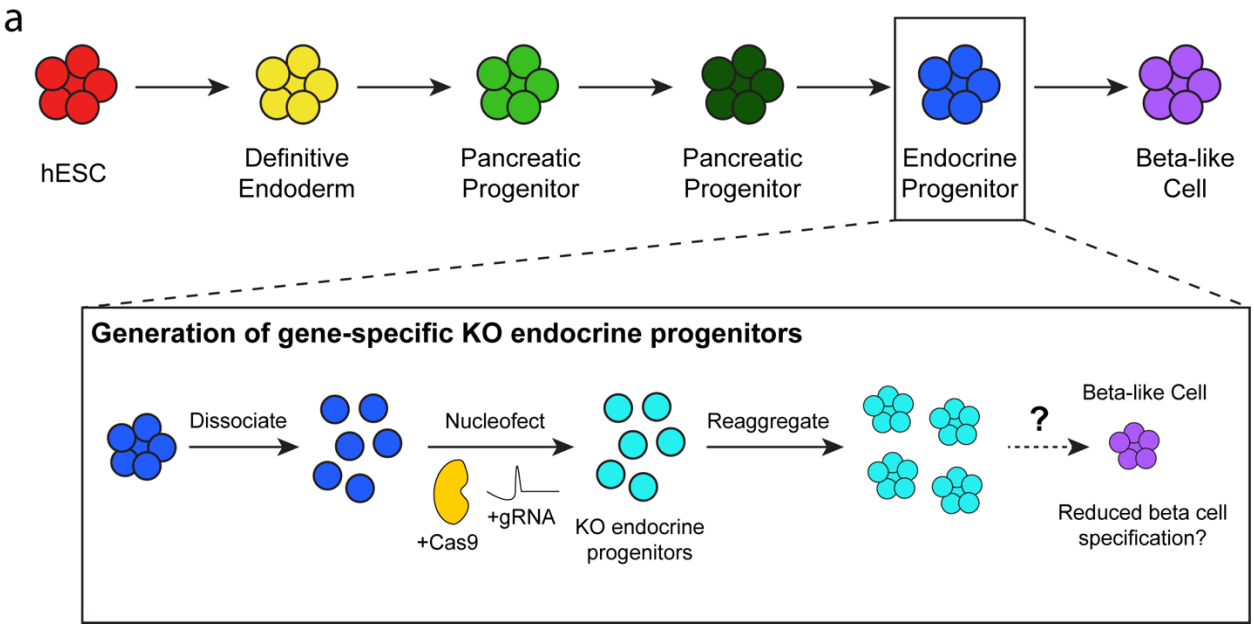


Figure 3.16. Development of a platform to functionally validate candidate beta lineage regulators. (a) Illustration depicting isolation of endocrine progenitor-stage cells for gene-knockdown via CRISPR/Cas9 genomic editing. Endocrine progenitor-stage clusters will be dissociated and nucleofected with Cas9 and a gRNA targeting a candidate beta lineage regulator. These edited endocrine progenitor-stage cells will then be re-aggregated and differentiated towards the beta lineage to determine if knockdown of specific candidate beta lineage regulators result in reduced beta cell differentiation.

Chapter 4

Conclusions and Future Directions

In this work, we have constructed a roadmap of murine pancreatic development by generating transcriptional profiles of known and novel cell populations in both the epithelial and mesenchymal compartments and inferring lineage relationships across developmental time (Chapter 2). Our discovery of a novel endocrine progenitor population defined by high *Fev* expression in mouse prompted us to examine human pancreatic development and uncover previously underappreciated heterogeneity in the endocrine compartment during differentiation (Chapter 3). We extended our findings from mouse to human and identified *FEV*-expressing endocrine progenitor populations in human pancreatic development. Using *in silico* modeling to construct lineage relationships, we mapped the transcriptional dynamics of human endocrine cell differentiation and identified candidate lineage regulators of human alpha and beta lineage allocation. We subsequently identified major differences present between human endocrine cell differentiation *in vivo* during development versus current *in vitro* methods to produce beta cells from hESCs. Our analyses identified a potential mis-differentiated lineage that formed as hESC-derived pancreatic progenitors differentiated towards a beta lineage. Blocking the generation of this mis-differentiated lineage during *in vitro* beta cell differentiation may be powerful approach to making *in vitro* beta cell differentiation more efficient and more reflective of *in vivo* beta cell development. Thus, our work forms a foundation on which improvements to *in vitro* beta cell differentiation can be made to more closely reflect proper endocrine cell development *in vivo*, to ideally increase both the yield and functionality of beta-like cells that are generated at the completion of the directed differentiation protocol.

Transcriptional stages and regulators of endocrine lineage allocation

How a seemingly homogenous endocrine progenitor population marked by *Ngn3* expression gives rise to a diversity of hormone-expressing lineages in the pancreas has been a subject of intense research. Our single-cell transcriptional dataset strengthened our

understanding of the events that occur following *Ngn3* downregulation and before endocrine cell specification in both mouse and human. In mouse, the time between *Ngn3* downregulation and hormone acquisition is reported to be only on the order of 12 hours (Beucher et al., 2012b). This short time window is also thought to be preserved in human cells but has not been experimentally determined (Jennings et al., 2015). It is possible that endocrine lineage commitment to a specific hormone-expressing cell type has already been made at the *Ngn3*-expressing stage and occurs prior to *Ngn3* downregulation. However, our analyses in both mouse and human pancreatic development suggested that commitment to a specific hormone lineage may occur at a stage following *Ngn3* expression. In our *in silico* reconstruction of endocrine cell differentiation, we observed *Fev/FEV*-expressing progenitors that appeared to become specified towards one hormone lineage over the other. This observation was not made in the common *Ngn3/NGN3*-expressing progenitor population present in both mouse and human, suggesting that commitment to a hormone lineage does not occur in this *Ngn3/NGN3*-expressing stage. One possible mechanism through which these pre-fated endocrine progenitor and differentiated hormone-expressing lineages are compartmentalized is via Wnt signaling. Inhibition of Wnt signaling promotes beta cell differentiation from hESC-derived endocrine progenitors *in vitro* (Sharon et al., 2019b). Wnt signaling may serve as a key regulator that ensures that post-Ngn3/NGN3+ endocrine progenitors have enough time to become specified towards a specific hormone-expressing lineage. Future work should focus on elucidating the precise cellular stages of endocrine lineage commitment.

Our work has uncovered candidate transcriptional regulators of alpha and beta lineage allocation in both murine and human endocrine cell development. Validation of their functional significance should be performed next. For candidate alpha lineage regulators, genetic knockout mice can be generated and studied for alpha cell differentiation defects. Mouse pancreatic explants also represents another method with which these candidate alpha regulators can be knocked down via viral transduction of gene-specific shRNA. This strategy can also be

leveraged in the directed differentiation of hESCs to glucagon-expressing alpha cells to validate the human alpha lineage regulators we have identified (Rezania et al., 2011). Similarly, our *in vitro* beta cell differentiation platform represents a method for validating the function of candidate beta lineage regulators. Using the CRISPR/Cas9-mediated gene knockdown system that we have developed, candidate beta lineage regulators can be specifically knocked down in endocrine progenitors to determine if differentiation to the beta lineage is impaired. Use of a human fetal pancreatic explant system may also represent an alternative system to functionally test these beta lineage regulators in a non-artificial, non-hESC-based system.

The role of the epigenome in endocrine lineage specification

Application of a recently developed technique called single-cell Assay for Transposase-Accessible Chromatin using sequencing (scATAC-seq) to endocrine progenitors could identify endocrine progenitor heterogeneity at the epigenetic level. Although transcriptionally similar, certain subsets of NGN3+ and FEV+ progenitors may differ dramatically in their potential to give rise to a specific hormone lineage. Specifically, are there NGN3+ or FEV+ progenitors present in pancreatic development that are differentially primed to express specific combinations of alpha or beta-specific genes? This technology can also be applied to PDX1+ pancreatic progenitors to determine if hormone-lineage specification extends even before the NGN3-expressing endocrine progenitor stage.

Function of *FEV* in human endocrine cell development

The transcription factor *FEV* was expressed at multiple stages of human endocrine cell development. *FEV* expression was one defining feature of pre-alpha and pre-beta progenitors found in our single-cell transcriptomic profiling of 12wpc pancreas. In our *in silico* reconstruction of lineage relationships, these *FEV*-expressing pre-alpha and pre-beta progenitors emerged

after the *NGN3*-expressing common progenitor population, which is concordant with *Fev* being a transcriptional target of *Ngn3* in mouse pancreatic development (Ohta et al., 2011). Unlike in the murine pancreas, however, *FEV* was only observed to be expressed in differentiated alpha cells and not beta cells in human fetal pancreas, suggesting that beta cells downregulate *FEV* upon differentiation into the beta lineage from pre-beta, *FEV*-expressing progenitors. In the human beta lineage, given that *FEV* was expressed in the pre-beta progenitor stage and appeared to be an important transcription factor for human beta cell differentiation, elucidation of its transcriptional targets will enhance our understanding of beta cell differentiation. This can be done either with human fetal tissue, using a FEV antibody suitable for ChIP-seq, or with cells from *in vitro* beta cell differentiation, leveraging the FEV-MYC hESC line whose construction we have described in this work. Identifying the transcriptional targets of FEV along the alpha lineage will also yield new insights into the transcriptional regulation of alpha cell differentiation. Identification of transcriptional targets along alpha cell differentiation and how they differ from those along beta cell differentiation will further our understanding of the transcriptional networks that regulate fate restriction of endocrine progenitors as they differentiate into distinct lineages. Given that *FEV* expression was retained in differentiated alpha cells, future work should also determine the role of *FEV* in alpha cell function.

Possible functions of *FEV* in human endocrine cell development may also be gleaned from known transcriptional targets of *Fev* in mouse serotonergic neurons and beta cells. In serotonergic neurons, Fev activates genes required for the acquisition of serotonin transmitter identity and for the functional maturation of serotonergic neurons into excitable cells in the brain (Wyler et al., 2016). In mouse insulinoma cells, Fev binds to the *Insulin1* and *Insulin2* promoters and regulates their transcription (Ohta et al., 2011). A comparison between Fev targets in the brain compared to those of the pancreas may offer insight into shared transcriptional circuitry that regulates both differentiation and function of serotonergic neurons and beta cells. These

insights can then be applied to future work investigating the targets of FEV along the process of human endocrine cell differentiation.

The role of the changing microenvironment during endocrine cell differentiation

In addition to intrinsic programs that dictate endocrine lineage allocation, the surrounding microenvironment may play a significant role in shaping these decisions. In both our murine and human studies, we have identified significant heterogeneity in multiple compartments of the surrounding microenvironment, including the mesenchyme, vasculature, nerves, and immune system. Several subpopulations within each of these compartments may influence endocrine cell development by promoting one hormone-expressing lineage over others. In mouse, we classified the developing mesenchyme into different cell populations across developmental time based on transcriptional heterogeneity. For instance, the mesothelium, a mesenchymal population within the developing pancreas, formed as the pancreatic epithelium developed. As the mesothelium matured, the profile of the factors it secreted changed, and the shift of mesothelium-derived secreted factors may provide critical cues that drive endocrine cell development. Similar to in mouse, several mesenchymal populations emerged in human pancreatic development, and the impact that these different mesenchymal cells have on the endocrine cell compartment is relatively unexplored in human pancreatic development. Multiple endothelial, neural, and immune populations also emerged across human development. Mapping how these non-epithelial populations change throughout the course of pancreatic development may yield new insight into how these populations influence endocrine cell development at discrete developmental times through secreted factors or direct physical interaction.

Improving strategies to generate beta-like cells *in vitro*

The ability to generate the insulin-producing, beta lineage from hESCs stems from knowledge of how beta cells form *in vivo*. With new insight gained from our mouse and human studies on endocrine cell development, we can implement modifications to the existing methods of *in vitro* beta cell differentiation. From our *in vivo* human fetal single-cell datasets, we can determine which transcriptional regulators are upregulated during proper beta cell differentiation. Following validation of these candidate beta lineage regulators, we can then determine if any of these transcriptional regulators fail to become expressed in existing *in vitro* beta cell differentiation protocols. If so, the addition of specific exogenous growth factors can be utilized to specifically activate gene expression of fundamental regulators of beta cell differentiation. Properly mimicking the transcriptional beta cell differentiation program may not only improve efficiency of beta cell differentiation *in vitro* but may also produce beta cells that display improved function and maintain this function over time.

Inhibition of mis-differentiation will also improve current methods of beta cell differentiation *in vitro*. In our *in vitro* beta cell differentiation platform, we uncovered a mis-differentiated lineage that resembled cells of the noradrenergic lineage defined by *PHOX2A* expression. Additionally, the generation of an enterochromaffin (EC) lineage from the same endocrine progenitor stage that gives rise to the beta lineage was observed (Veres et al., 2019). Effectively suppressing the emergence of these lineages during the process of generating beta cells from hESCs is expected to improve yield outcomes, which is an important consideration when there is a need to generate enough beta cells for eventual human clinical transplantation. For instance, use of BMP and cAMP inhibitors may prevent the emergence of the PHOX2A+ noradrenergic lineage, and use of small molecules that suppress EC differentiation may inhibit the formation of ECs during beta cell differentiation.

Finally, since beta cell development does not occur in isolation, the addition of specific cell types and/or their secreted factors to defined stages along beta cell differentiation may also

improve efficiency of *in vitro* beta cell differentiation or function of the resulting hESC-derived beta cells. For instance, the pancreatic or endocrine progenitor cell stages of *in vitro* beta cell differentiation may benefit from the additional of specific mesenchymal, endothelial, or nerve cells that are observed to interact with endocrine progenitors *in vivo*. With the transcriptomic data we have at single-cell resolution of these populations present through human pancreatic development, future work can isolate specific populations and perform co-culture studies in conjunction with *in vitro* beta cell differentiation. Isolation of these cells can also enable generation of primary cell lines that can be stably maintained in culture for use in co-culture studies. Alternatively, deeper analysis of the transcriptomic profiles of these niche populations present in human endocrine cell development can reveal specific combinations of secreted factors that can be added *in vitro* to support beta cell development in lieu of adding the niche cells themselves.

In vitro generation of other pancreatic hormone-expressing lineages from FEV+ progenitors

Although generating beta cells from hESCs is a main focus of the field given their direct relevance to diabetes, we must also think beyond the beta cell and towards other islet lineages if we wish to assemble hESC-derived beta cells into pseudo-islets that resemble islets *in vivo*. Our murine study of endocrine cell development revealed that the alpha, beta, delta, epsilon, and gamma lineages of the pancreas transit through a Fev-expressing cell stage, likely through a Fev+ endocrine progenitor stage. Through examining human endocrine cell development, we observed that both alpha and beta lineages also are derived from a FEV+ progenitor stage. Higher resolution of human endocrine cell differentiation will benefit from increased numbers of the delta, epsilon, and gamma cells. Finally, in our *in vitro* model of beta cell differentiation, cells that express SST also appear to transit through a FEV-expressing cell stage, suggesting that

human differentiation into at least SST-expressing cells involves transit through a FEV+ progenitor.

Given our findings in murine and human endocrine cell differentiation, it is possible that FEV+ progenitors serve as a common endocrine stage that precedes differentiation into all five hormone-expressing islet lineages in human. Additional single-cell RNA-sequencing of cells enriched in the endocrine lineage will increase the numbers of delta, epsilon, and gamma cells that will enable *in silico* lineage tracing for these lineages. This additional transcriptomic data can also identify candidate regulators of delta, epsilon, and gamma lineages. The transcriptional drivers of these endocrine lineages have been largely unexplored in both mouse and human, as efforts have been mainly focused on the alpha and beta lineages. The FEV+ endocrine progenitors we identify in *in vitro* beta cell differentiation may serve as a launch point for generating other non-beta endocrine lineages *in vitro*. Future work can leverage our FEV reporter lines, enrich for FEV+ cells at the endocrine progenitor cell stage, and screen for the generation of non-beta endocrine lineages using different cocktails of exogenous growth factors.

Modeling human pancreatic development in a dish

This work greatly augments our ability to model human pancreatic development *in vitro*. By applying single-cell RNA-sequencing to both mouse and human pancreatic development, we have uncovered novel stages of endocrine cell development that better inform our goal of differentiating beta cells from hESCs. Single-cell RNA-sequencing is a valuable tool to learn new differentiation trajectories in which genetic lineage tracing *in vivo* is not possible, such as in human fetal development. Applying this tool also enabled the identification of a mis-differentiated lineage that emerged during *in vitro* beta cell differentiation. This mis-differentiated lineage greatly diminishes the efficiency of generating hESC-derived beta cells. By combining the findings from our single-cell sequencing of murine and human pancreatic development and of *in vitro* beta cell differentiation, we can now begin implementing improvements to the current

process of the directed differentiation of beta cells such that we can soon generate functional beta cells more efficiently for cellular replacement therapy for individuals with diabetes mellitus.

References

Ahlgren, U., Pfaff, S.L., Jessell, T.M., Edlund, T., and Edlund, H. (1997). Notochord to endoderm signaling is required for pancreas development. Nature *124*, 4243–4252.

Ahnfelt-Rønne, J., Ravassard, P., Pardanaud-Glavieux, C., Scharfmann, R., and Serup, P. (2010). Mesenchymal bone morphogenetic protein signaling is required for normal pancreas development. Diabetes *59*, 1948–1956.

Alvarez, E., Zhou, W., Witta, S.E., and Freed, C.R. (2005). Characterization of the Bex gene family in humans, mice, and rats. Gene *357*, 18–28.

Angelo, J.R., and Tremblay, K.D. (2018). Identification and fate mapping of the pancreatic mesenchyme. Dev. Biol. *435*, 15–25.

Apelqvist, A., Li, H., Sommer, L., Beatus, P., Anderson, D.J., Honjo, T., Hrabe de Angelis, M., Lendahl, U., and Edlund, H. (1999). Notch signalling controls pancreatic cell differentiation. Nature *400*, 877–881.

Arnes, L., Hill, J.T., Gross, S., Magnuson, M.A., and Sussel, L. (2012). Ghrelin Expression in the Mouse Pancreas Defines a Unique Multipotent Progenitor Population. PLoS ONE *7*, e52026.

Artner, I., Blanchi, B., Raum, J.C., Guo, M., Kaneko, T., Cordes, S., Sieweke, M., and Stein, R. (2007). MafB is required for islet beta cell maturation. Proc. Natl. Acad. Sci. U.S.a. *104*, 3853–3858.

Artner, I., Hang, Y., Mazur, M., Yamamoto, T., Guo, M., Lindner, J., Magnuson, M.A., and Stein, R. (2010). MafA and MafB regulate genes critical to beta-cells in a unique temporal manner. Diabetes *59*, 2530–2539.

Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., Ryu, J.H., Wagner,

B.K., Shen-Orr, S.S., Klein, A.M., et al. (2016). A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. Cell Syst *3*, 346–360.e4.

Basta, G., Montanucci, P., Luca, G., Boselli, C., Noya, G., Barbaro, B., Qi, M., Kinzer, K.P., Oberholzer, J., and Calafiore, R. (2011). Long-term metabolic and immunological follow-up of nonimmunosuppressed patients with type 1 diabetes treated with microencapsulated islet allografts: four cases. Diabetes Care *34*, 2406–2409.

Bedoyan, J.K., Kumar, R.A., Sudi, J., Silverstein, F., Ackley, T., Iyer, R.K., Christian, S.L., and Martin, D.M. (2010). Duplication 16p11.2 in a child with infantile seizure disorder. Am. J. Med. Genet. A *152A*, 1567–1574.

Benitez, C.M., Qu, K., Sugiyama, T., Pauerstein, P.T., Liu, Y., Tsai, J., Gu, X., Ghodasara, A., Arda, H.E., Zhang, J., et al. (2014). An integrated cell purification and genomics strategy reveals multiple regulators of pancreas development. PLoS Genet. *10*, e1004645.

Benjanirut, C., Paris, M., Wang, W.-H., Hong, S.J., Kim, K.S., Hullinger, R.L., and Andrisani, O.M. (2006). The cAMP pathway in combination with BMP2 regulates Phox2a transcription via cAMP response element binding sites. Journal of Biological Chemistry *281*, 2969–2981.

Bertolino, P., Holmberg, R., Reissmann, E., Andersson, O., Berggren, P.-O., and Ibáñez, C.F. (2008). Activin B receptor ALK7 is a negative regulator of pancreatic beta-cell function. Proc. Natl. Acad. Sci. U.S.a. *105*, 7246–7251.

Beucher, A., Gjernes, E., Collin, C., Courtney, M., Meunier, A., Collombat, P., and Gradwohl, G. (2012a). The homeodomain-containing transcription factors Arx and Pax4 control enteroendocrine subtype specification in mice. PLoS ONE *7*, e36449.

Beucher, A., Martín, M., Spenle, C., Poulet, M., Collin, C., and Gradwohl, G. (2012b). Competence of failed endocrine progenitors to give rise to acinar but not ductal cells is restricted to early pancreas development. Dev. Biol. *361*, 277–285.

Bhushan, A., Itoh, N., Kato, S., Thiery, J.P., Czernichow, P., Bellusci, S., and Scharfmann, R. (2001). Fgf10 is essential for maintaining the proliferative capacity of epithelial progenitor cells during early pancreatic organogenesis. Development *128*, 5109–5117.

Bin Zhou, Honor, L.B., He, H., Ma, Q., Oh, J.-H., Butterfield, C., Lin, R.-Z., Melero-Martin, J.M., Dolmatova, E., Duffy, H.S., et al. (2011). Adult mouse epicardium modulates myocardial injury by secreting paracrine factors. J. Clin. Invest. *121*, 1894–1904.

Blake, J.A., and Ziman, M.R. (2014). Pax genes: regulators of lineage specification and progenitor cell maintenance. Development *141*, 737–751.

Blum, B., Hrvatin, S.S.Š., Schuetz, C., Bonal, C., Rezania, A., and Melton, D.A. (2012). Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3. Nat. Biotechnol. *30*, 261–264.

Bonini, C., Grez, M., Traversari, C., Ciceri, F., Marktel, S., Ferrari, G., Dinauer, M., Sadat, M., Aiuti, A., Deola, S., et al. (2003). Safety of retroviral gene marking with a truncated NGF receptor. Nature Medicine *9*, 367–369.

Bonnefond, A., Vaillant, E., Philippe, J., Skrobek, B., Lobbens, S., Yengo, L., Huyvaert, M., Cavé, H., Busiah, K., Scharfmann, R., et al. (2013). Transcription factor gene MNX1 is a novel cause of permanent neonatal diabetes in a consanguineous family. Diabetes Metab. *39*, 276–280.

Bonner-Weir, S., Aguayo-Mazzucato, C., and Weir, G.C. (2016). Dynamic development of the

pancreas from birth to adulthood. Ups. J. Med. Sci. *121*, 155–158.

Borden, P., Houtz, J., Leach, S.D., and Kuruvilla, R. (2013). Sympathetic innervation during development is necessary for pancreatic islet architecture and functional maturation. Cell Rep *4*, 287–301.

Borghini, S., Bachetti, T., Fava, M., Di Duca, M., Cargnin, F., Fornasari, D., Ravazzolo, R., and Ceccherini, I. (2006). The TLX2 homeobox gene is a transcriptional target of PHOX2B in neural-crest-derived cells. Biochem. J. *395*, 355–361.

Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. *42*, e168–e168.

Bruin, J.E., Asadi, A., Fox, J.K., Erener, S., Rezania, A., and Kieffer, T.J. (2015). Accelerated Maturation of Human Stem Cell-Derived Pancreatic Progenitor Cells into Insulin-Secreting Cells in Immunodeficient Rats Relative to Mice. Stem Cell Reports *5*, 1081–1096.

Butler, A.E., Dhawan, S., Hoang, J., Cory, M., Zeng, K., Fritsch, H., Meier, J.J., Rizza, R.A., and Butler, P.C. (2016). β-Cell Deficit in Obese Type 2 Diabetes, a Minor Role of β-Cell Dedifferentiation and Degranulation. The Journal of Clinical Endocrinology & Metabolism *101*, 523–532.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Byrnes, L.E., Wong, D.M., Subramaniam, M., Meyer, N.P., Gilchrist, C.L., Knox, S.M., Tward, A.D., Ye, C.J., and Sneddon, J.B. (2018). Lineage dynamics of murine pancreatic development

at single-cell resolution. Nat Commun *9*, 3922.

Cao, H., Jheon, A., Li, X., Sun, Z., Wang, J., Florez, S., Zhang, Z., McManus, M.T., Klein, O.D., and Amendt, B.A. (2013). The Pitx2:miR-200c/141:noggin pathway regulates Bmp signaling and ameloblast differentiation. Development *140*, 3348–3359.

Chen, S., Ji, M., Paris, M., Hullinger, R.L., and Andrisani, O.M. (2005). The cAMP pathway regulates both transcription and activity of the paired homeobox transcription factor Phox2a required for development of neural crest-derived and central nervous system-derived catecholaminergic neurons. Journal of Biological Chemistry *280*, 41025–41036.

Chiellini, C., Grenningloh, G., Cochet, O., Scheideler, M., Trajanoski, Z., Ailhaud, G., Dani, C., and Amri, E.-Z. (2008). Stathmin-like 2, a developmentally-associated neuronal marker, is expressed and modulated during osteogenesis of human mesenchymal stem cells. Biochemical and Biophysical Research Communications *374*, 64–68.

Cho, G.-S., Park, D.-S., Choi, S.-C., and Han, J.-K. (2017). Tbx2 regulates anterior neural specification by repressing FGF signaling pathway. Dev. Biol. *421*, 183–193.

Churchill, A.J., Gutiérrez, G.D., Singer, R.A., Lorberbaum, D.S., Fischer, K.A., and Sussel, L. (2017). Genetic evidence that Nkx2.2 acts primarily downstream of Neurog3 in pancreatic endocrine lineage development. eLife *6*, R106.

Collombat, P., Hecksher-Sørensen, J., Krull, J., Berger, J., Riedel, D., Herrera, P.L., Serup, P., and Mansouri, A. (2007). Embryonic endocrine pancreas and mature beta cells acquire alpha and PP cell phenotypes upon Arx misexpression. J. Clin. Invest. *117*, 961–970.

Collombat, P., Mansouri, A., Hecksher-Sørensen, J., Serup, P., Krull, J., Gradwohl, G., and Gruss, P. (2003). Opposing actions of Arx and Pax4 in endocrine pancreas development.

Genes Dev. *17*, 2591–2603.

Collombat, P., Xu, X., Ravassard, P., Sosa-Pineda, B., Dussaud, S.B., Billestrup, N., Madsen, O.D., Serup, P., Heimberg, H., and Mansouri, A. (2009). The Ectopic Expression of Pax4 in the Mouse Pancreas Converts Progenitor Cells into ? and Subsequently ? Cells. Cell *138*, 449–462.

Côté, F., Thévenot, E., Fligny, C., Fromes, Y., Darmon, M., Ripoche, M.-A., Bayard, E., Hanoun, N., Saurini, F., Lechat, P., et al. (2003). Disruption of the nonneuronal tph1 gene demonstrates the importance of peripheral serotonin in cardiac function. Pnas *100*, 13525–13530.

D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat. Biotechnol. *23*, 1534–1541.

D'Amour, K.A., Bang, A.G., Eliazer, S., Kelly, O.G., Agulnick, A.D., Smart, N.G., Moorman, M.A., Kroon, E., Carpenter, M.K., and Baetge, E.E. (2006). Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. Nat. Biotechnol. *24*, 1392–1401.

Da Silva Xavier, G. (2018). The Cells of the Islets of Langerhans. J Clin Med *7*, 54.

Dekel, B., Metsuyanim, S., Schmidt-Ott, K.M., Fridman, E., Jacob-Hirsch, J., Simon, A., Pinthus, J., Mor, Y., Barasch, J., Amariglio, N., et al. (2006). Multiple Imprinted and Stemness Genes Provide a Link between Normal and Tumor Progenitor Cells of the Developing Human Kidney. Cancer Res *66*, 6040–6049.

Desgraz, R., and Herrera, P.L. (2009). Pancreatic neurogenin 3-expressing cells are unipotent islet precursors. Development *136*, 3567–3574.

Dever, D.P., Bak, R.O., Reinisch, A., Camarena, J., Washington, G., Nicolas, C.E., Pavel-Dinu, M., Saxena, N., Wilkens, A.B., Mantri, S., et al. (2016). CRISPR/Cas9 β-globin gene targeting in

human haematopoietic stem cells. Nature *539*, 384–389.

Ding, Y.-Q., Marklund, U., Yuan, W., Yin, J., Wegman, L., Ericson, J., Deneris, E., Johnson, R.L., and Chen, Z.-F. (2003). Lmx1b is essential for the development of serotonergic neurons. Nat. Neurosci. *6*, 933–938.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2012). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Doering, J.E., Kane, K., Hsiao, Y.-C., Yao, C., Shi, B., Slowik, A.D., Dhagat, B., Scott, D.D., Ault, J.G., Page-McCaw, P.S., et al. (2008). Species differences in the expression of Ahi1, a protein implicated in the neurodevelopmental disorder Joubert syndrome, with preferential accumulation to stigmoid bodies. J. Comp. Neurol. *511*, 238–256.

Dong, H., Zhang, Y., Wang, J., Kim, D.S., Wu, H., Sjögren, B., Gao, W., Luttrell, L., and Wang, H. (2017). Regulator of G protein signaling 2 is a key regulator of pancreatic β-cell mass and function. Cell Death Dis *8*, e2821–e2821.

Dorajoo, R., Ali, Y., Tay, V.S.Y., Kang, J., Samydurai, S., Liu, J., and Boehm, B.O. (2017). Single-cell transcriptomics of East-Asian pancreatic islets cells. Sci Rep *7*, 5024.

Dorrell, C., Schug, J., Canaday, P.S., Russ, H.A., Tarlow, B.D., Grompe, M.T., Horton, T., Hebrok, M., Streeter, P.R., Kaestner, K.H., et al. (2016). Human islets contain four distinct subtypes of β cells. Nat Commun *7*, 11756.

Du, A., Hunter, C.S., Murray, J., Noble, D., Cai, C.-L., Evans, S.M., Stein, R., and May, C.L. (2009). Islet-1 is required for the maturation, proliferation, and survival of the endocrine pancreas. Diabetes *58*, 2059–2069.

Ernst, M.C., and Sinal, C.J. (2010). Chemerin: at the crossroads of inflammation and obesity. Trends in Endocrinology & Metabolism *21*, 660–667.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. *16*, 278.

Flanagan, S.E., De Franco, E., Lango Allen, H., Zerah, M., Abdul-Rasoul, M.M., Edge, J.A., Stewart, H., Alamiri, E., Hussain, K., Wallis, S., et al. (2014). Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. Cell Metab. *19*, 146–154.

Frank, P.G. (2003). Caveolin, Caveolae, and Endothelial Cell Function. Arteriosclerosis, Thrombosis, and Vascular Biology *23*, 1161–1168.

Ge, W., He, F., Kim, K.J., Blanchi, B., Coskun, V., Nguyen, L., Wu, X., Zhao, J., Heng, J.I.-T., Martinowich, K., et al. (2006). Coupling of cell migration with neurogenesis by proneural bHLH factors. Pnas *103*, 1319–1324.

Gehart, H., van Es, J.H., Hamer, K., Beumer, J., Kretzschmar, K., Dekkers, J.F., Rios, A., and Clevers, H. (2019). Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping. Cell *176*, 1158–1173.e16.

Gershon, M.D. (2013). 5-Hydroxytryptamine (serotonin) in the gastrointestinal tract. Curr Opin Endocrinol Diabetes Obes *20*, 14–21.

Gierl, M.S., Karoulias, N., Wende, H., Strehle, M., and Birchmeier, C. (2006). The zinc-finger factor Insm1 (IA-1) is essential for the development of pancreatic beta cells and intestinal

endocrine cells. Genes Dev. *20*, 2465–2478.

Gilchrist, J.A., Best, C.H., and Banting, F.G. (1923). Observations with Insulin on Department of Soldiers' Civil Re-Establishment Diabetics. Can Med Assoc J *13*, 565–572.

Giorgakis, E., Mathur, A.K., Chakkera, H.A., Reddy, K.S., Moss, A.A., and Singer, A.L. (2018). Solid pancreas transplant: Pushing forward. World J Transplant *8*, 237–251.

Golosow, N., and Grobstein, C. (1962). Epitheliomesenchymal interaction in pancreatic morphogenesis. Dev. Biol. *4*, 242–255.

Gouzi, M., Kim, Y.H., Katsumoto, K., Johansson, K., and Grapin-Botton, A. (2011). Neurogenin3 initiates stepwise delamination of differentiating endocrine cells during pancreas development. Dev. Dyn. *240*, 589–604.

Gradwohl, G., Dierich, A., LeMeur, M., and Guillemot, F. (2000). neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. Proc. Natl. Acad. Sci. U.S.a. *97*, 1607–1611.

Grenningloh, G., Soehrman, S., Bondallaz, P., Ruchti, E., and Cadas, H. (2003). Role of the microtubule destabilizing proteins SCG10 and stathmin in neuronal growth. Journal of Neurobiology *58*, 60–69.

Gross, S., Garofalo, D.C., Balderes, D.A., Mastracci, T.L., Dias, J.M., Perlmann, T., Ericson, J., and Sussel, L. (2016). The novel enterochromaffin marker Lmx1a regulates serotonin biosynthesis in enteroendocrine cell lineages downstream of Nkx2.2. Development *143*, 2616–2628.

Gu, G., Brown, J.R., and Melton, D.A. (2003). Direct lineage tracing reveals the ontogeny of pancreatic cell fates during mouse embryogenesis. Mechanisms of Development *120*, 35–43.

Gu, G., Dubauskaite, J., and Melton, D.A. (2002). Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. Development *129*, 2447–2457.

Guillemot, F., Lo, L.C., Johnson, J.E., Auerbach, A., Anderson, D.J., and Joyner, A.L. (1993). Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. Cell *75*, 463–476.

Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427.

Hald, J., Galbo, T., Rescan, C., Radzikowski, L., Sprinkel, A.E., Heimberg, H., Ahnfelt-Rønne, J., Jensen, J., Scharfmann, R., Gradwohl, G., et al. (2011). Pancreatic islet and progenitor cell surface markers with cell sorting potential. Diabetologia *55*, 154–165.

Han, S.-I., Yasuda, K., and Kataoka, K. (2011). ATF2 interacts with beta-cell-enriched transcription factors, MafA, Pdx1, and beta2, and activates insulin gene transcription. Journal of Biological Chemistry *286*, 10449–10456.

Hang, Y., and Stein, R. (2011). MafA and MafB activity in pancreatic β cells. Trends Endocrinol. Metab. *22*, 364–373.

Hasegawa, M., Higashi, K., Matsushita, T., Hamaguchi, Y., Saito, K., Fujimoto, M., and Takehara, K. (2013). Dermokine inhibits ELR+CXC chemokine expression and delays early skin wound healing. Journal of Dermatological Science *70*, 34–41.

Haugas, M., Tikker, L., Achim, K., Salminen, M., and Partanen, J. (2016). Gata2 and Gata3 regulate the differentiation of serotonergic and glutamatergic neuron subtypes of the dorsal

raphe. Development *143*, 4495–4508.

Haumaitre, C., Barbacci, E., Jenny, M., Ott, M.O., Gradwohl, G., and Cereghini, S. (2005). Lack of TCF2/vHNF1 in mice leads to pancreas agenesis. Pnas *102*, 1490–1495.

Hay, E.D. (2005). The mesenchymal cell, its role in the embryo, and the remarkable signaling mechanisms that create it. Dev. Dyn. *233*, 706–720.

Hayashida, K.-I., and Eisenach, J.C. (2018). Descending Noradrenergic Inhibition: An Important Mechanism of Gabapentin Analgesia in Neuropathic Pain. Adv. Exp. Med. Biol. *1099*, 93–100.

Hayden, M.R., Patel, K., Habibi, J., Gupta, D., Tekwani, S.S., Whaley-Connell, A., and Sowers, J.R. (2008). Attenuation of endocrine-exocrine pancreatic communication in type 2 diabetes: pancreatic extracellular matrix ultrastructural abnormalities. J Cardiometab Syndr *3*, 234–243.

Hebrok, M., Kim, S.K., and Melton, D.A. (1998). Notochord repression of endodermal Sonic hedgehog permits pancreas development. Genes Dev. *12*, 1705–1713.

Hecksher-Sørensen, J., Watson, R.P., Lettice, L.A., Serup, P., Eley, L., De Angelis, C., Ahlgren, U., and Hill, R.E. (2004). The splanchnic mesodermal plate directs spleen and pancreatic laterality, and is regulated by Bapx1/Nkx3.2. Development *131*, 4665–4675.

Heller, C., Kühn, M.C., Mülders-Opgenoorth, B., Schott, M., Willenberg, H.S., Scherbaum, W.A., and Schinner, S. (2011). Exendin-4 upregulates the expression of Wnt-4, a novel regulator of pancreatic β-cell proliferation. Am. J. Physiol. Endocrinol. Metab. *301*, E864–E872.

Heller, R.S., Jenny, M., Collombat, P., Mansouri, A., Tomasetto, C., Madsen, O.D., Mellitzer, G., Gradwohl, G., and Serup, P. (2005). Genetic determinants of pancreatic ε-cell development. Dev. Biol. *286*, 217–224.

Hendricks, T., Francis, N., Fyodorov, D., and Deneris, E.S. (1999). The ETS domain factor Pet-1 is an early and precise marker of central serotonin neurons and interacts with a conserved element in serotonergic genes. J. Neurosci. *19*, 10348–10356.

Hernandez-Torres, F., Rodríguez-Outeiriño, L., Franco, D., and Aranega, A.E. (2017). Pitx2 in Embryonic and Adult Myogenesis. Front. Cell Dev. Biol. *5*, 211.

Herrera, P.L. (2000). Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. Development *127*, 2317–2322.

Herrera, P.L., Orci, L., and Vassalli, J.D. (1998). Two transgenic approaches to define the cell lineages in endocrine pancreas development. Mol. Cell. Endocrinol. *140*, 45–50.

Hirsch, M.R., Tiveron, M.C., Guillemot, F., Brunet, J.F., and Goridis, C. (1998). Control of noradrenergic differentiation and Phox2a expression by MASH1 in the central and peripheral nervous system. Development *125*, 599–608.

Hishida, T., Naito, K., Osada, S., Nishizuka, M., and Imagawa, M. (2007). peg10, an imprinted gene, plays a crucial role in adipocyte differentiation. FEBS Lett. *581*, 4272–4278.

Hoffmann, A., and Spengler, D. (2012). Transient neonatal diabetes mellitus gene Zac1 impairs insulin secretion in mice through Rasgrf1. Mol. Cell. Biol. *32*, 2549–2560.

Hori, K., Nagai, T., Shan, W., Sakamoto, A., Taya, S., Hashimoto, R., Hayashi, T., Abe, M., Yamazaki, M., Nakao, K., et al. (2014). Cytoskeletal Regulation by AUTS2 in Neuronal Migration and Neuritogenesis. CellReports *9*, 2166–2179.

Huang, C., Xiang, Y., Chen, S., Yu, H., Wen, Z., Ye, T., Sun, H., Kong, H., Li, D., Yu, D., et al. (2017). Dermokine contributes to epithelial–mesenchymal transition through increased activation of signal transducer and activator of transcription 3 in pancreatic cancer. Cancer

Science *108*, 2130–2141.

Huang, E.S., Brown, S.E.S., Ewigman, B.G., Foley, E.C., and Meltzer, D.O. (2007). Patient perceptions of quality of life with diabetes-related complications and treatments. Diabetes Care *30*, 2478–2483.

Huang, H.P., Liu, M., El-Hodiri, H.M., Chu, K., Jamrich, M., and Tsai, M.J. (2000). Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3. Mol. Cell. Biol. *20*, 3292–3307.

Jacobson, E.F., and Tzanakakis, E.S. (2017). Human pluripotent stem cell differentiation to functional pancreatic cells for diabetes therapies: Innovations, challenges and future directions. J Biol Eng *11*, 21.

Jayewickreme, C.D., and Shivdasani, R.A. (2015). Control of stomach smooth muscle development and intestinal rotation by transcription factor BARX1. Dev. Biol. *405*, 21–32.

Jennings, R.E., Berry, A.A., Kirkwood-Wilson, R., Roberts, N.A., Hearn, T., Salisbury, R.J., Blaylock, J., Piper Hanley, K., and Hanley, N.A. (2013). Development of the human pancreas from foregut to endocrine commitment. Diabetes *62*, 3514–3522.

Jennings, R.E., Berry, A.A., Strutt, J.P., Gerrard, D.T., and Hanley, N.A. (2015). Human pancreas development. Development *142*, 3126–3137.

Jenny, M., Uhl, C., Roche, C., Duluc, I., Guillermin, V., Guillemot, F., Jensen, J., Kedinger, M., and Gradwohl, G. (2002). Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium. Embo J. *21*, 6338–6347.

Jensen, J.N., Cameron, E., Garay, M.V.R., Starkey, T.W., Gianani, R., and Jensen, J. (2005). Recapitulation of elements of embryonic development in adult mouse pancreatic regeneration.

Gastroenterology *128*, 728–741.

Jeon, J., Correa-Medina, M., Ricordi, C., Edlund, H., and Diez, J.A. (2009). Endocrine cell clustering during human pancreas development. J. Histochem. Cytochem. *57*, 811–824.

Johansson, K.A., Dursun, U., Jordan, N., Gu, G., Beermann, F., Gradwohl, G., and Grapin-Botton, A. (2007). Temporal control of neurogenin3 activity in pancreas progenitors reveals competence windows for the generation of different endocrine cell types. Dev. Cell *12*, 457–465.

Johnston, N.R., Mitchell, R.K., Haythorne, E., Pessoa, M.P., Semplici, F., Ferrer, J., Piemonti, L., Marchetti, P., Bugliani, M., Bosco, D., et al. (2016). Beta Cell Hubs Dictate Pancreatic Islet Responses to Glucose. Cell Metab. *24*, 389–401.

Kamburov, A., Wierling, C., Lehrach, H., and Herwig, R. (2008). ConsensusPathDB—a database for integrating human functional interaction networks. Nucleic Acids Res. *37*, D623–D628.

Kameswaran, V., Bramswig, N.C., McKenna, L.B., Penn, M., Schug, J., Hand, N.J., Chen, Y., Choi, I., Vourekas, A., Won, K.-J., et al. (2014). Epigenetic regulation of the DLK1-MEG3 microRNA cluster in human type 2 diabetic islets. Cell Metab. *19*, 135–145.

Kamiya, M., Judson, H., Okazaki, Y., Kusakabe, M., Muramatsu, M., Takada, S., Takagi, N., Arima, T., Wake, N., Kamimura, K., et al. (2000). The cell cycle control gene ZAC/PLAGL1 is imprinted--a strong candidate gene for transient neonatal diabetes. Hum. Mol. Genet. *9*, 453–460.

Kanamori-Katayama, M., Kaiho, A., Ishizu, Y., Okamura-Oho, Y., Hino, O., Abe, M., Kishimoto, T., Sekihara, H., Nakamura, Y., Suzuki, H., et al. (2011). LRRN4 and UPK3B are markers of primary mesothelial cells. PLoS ONE *6*, e25391.

Kapadia, C., Ghosh, M.C., Grass, L., and Diamandis, E.P. (2004). Human kallikrein 13 involvement in extracellular matrix degradation. Biochemical and Biophysical Research Communications *323*, 1084–1090.

Kelly, W.D., Lillehei, R.C., Merkel, F.K., Idezuki, Y., and Goetz, F.C. (1967). Allotransplantation of the pancreas and duodenum along with the kidney in diabetic nephropathy. Surgery *61*, 827–837.

Kesavan, G., Sand, F.W., Greiner, T.U., Johansson, J.K., Kobberup, S., Wu, X., Brakebusch, C., and Semb, H. (2009). Cdc42-mediated tubulogenesis controls cell specification. Cell *139*, 791–801.

Kim, A., Miller, K., Jo, J., Kilimnik, G., Wojcik, P., and Hara, M. (2009). Islet architecture: A comparative study. Islets *1*, 129–136.

Kiyasova, V., and Gaspar, P. (2011). Development of raphe serotonin neurons from specification to guidance. Eur. J. Neurosci. *34*, 1553–1562.

Klim, J.R., Williams, L.A., Limone, F., Guerra San Juan, I., Davis-Dusenbery, B.N., Mordes, D.A., Burberry, A., Steinbaugh, M.J., Gamage, K.K., Kirchner, R., et al. (2019). ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair. Nat. Neurosci. *22*, 167–179.

Krentz, N.A.J., Lee, M.Y.Y., Xu, E.E., Sproul, S.L.J., Maslova, A., Sasaki, S., and Lynn, F.C. (2018). Single-Cell Transcriptome Profiling of Mouse and hESC-Derived Pancreatic Progenitors. Stem Cell Reports *11*, 1551–1564.

Küry, P., Greiner-Petter, R., Cornely, C., Jürgens, T., and Müller, H.W. (2002). Mammalian achaete scute homolog 2 is expressed in the adult sciatic nerve and regulates the expression of

Krox24, Mob-1, CXCR4, and p57kip2 in Schwann cells. J. Neurosci. *22*, 7586–7595.

Kwapiszewska, G., Wygrecka, M., Marsh, L.M., Schmitt, S., Trosser, R., Wilhelm, J., Helmus, K., Eul, B., Zakrzewicz, A., Ghofrani, H.A., et al. (2008). Fhl-1, a New Key Protein in Pulmonary Hypertension. Circulation *118*, 1183–1194.

Lammert, E., Cleaver, O., and Melton, D. (2001). Induction of pancreatic differentiation by signals from blood vessels. Science *294*, 564–567.

Lammert, E., Cleaver, O., and Melton, D. (2003). Role of endothelial cells in early pancreas and liver development. Mechanisms of Development *120*, 59–64.

Landsman, L., Nijagal, A., Whitchurch, T.J., Vanderlaan, R.L., Zimmer, W.E., Mackenzie, T.C., and Hebrok, M. (2011). Pancreatic mesenchyme regulates epithelial organogenesis throughout development. PLoS Biol. *9*, e1001143.

Larsen, B.M., Hrycaj, S.M., Newman, M., Li, Y., and Wellik, D.M. (2015). Mesenchymal Hox6function is required for mouse pancreatic endocrine cell differentiation. Development *142*, 3859–3868.

Lau, J., Kawahira, H., and Hebrok, M. (2006). Hedgehog signaling in pancreas development and disease. Cell. Mol. Life Sci. *63*, 642–652.

Lawlor, N., George, J., Bolisetty, M., Kursawe, R., Sun, L., Sivakamasundari, V., Kycia, I., Robson, P., and Stitzel, M.L. (2017). Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. Genome Res. *27*, 208–222.

Liu, C., Maejima, T., Wyler, S.C., Casadesus, G., Herlitze, S., and Deneris, E.S. (2010). Pet-1 is required across different stages of life to regulate serotonergic function. Nat. Neurosci. *13*,

1190–1198.

Lo, L., Morin, X., Brunet, J.F., and Anderson, D.J. (1999). Specification of neurotransmitter identity by Phox2 proteins in neural crest stem cells. Neuron *22*, 693–705.

Lo, L., Tiveron, M.C., and Anderson, D.J. (1998). MASH1 activates expression of the paired homeodomain transcription factor Phox2a, and couples pan-neuronal and subtype-specific components of autonomic neuronal identity. Development *125*, 609–620.

Lock, L.T., and Tzanakakis, E.S. (2007). Stem/Progenitor cell sources of insulin-producing cells for the treatment of diabetes. Tissue Eng. *13*, 1399–1412.

López-Díaz, L., Jain, R.N., Keeley, T.M., VanDussen, K.L., Brunkan, C.S., Gumucio, D.L., and Samuelson, L.C. (2007). Intestinal Neurogenin 3 directs differentiation of a bipotential secretory progenitor to endocrine cell rather than goblet cell fate. Dev. Biol. *309*, 298–305.

Lund, M.L., Egerod, K.L., Engelstoft, M.S., Dmytriyeva, O., Theodorsson, E., Patel, B.A., and Schwartz, T.W. (2018). Enterochromaffin 5-HT cells - A major target for GLP-1 and gut microbial metabolites. Mol Metab *11*, 70–83.

Lyttle, B.M., Li, J., Krishnamurthy, M., Fellows, F., Wheeler, M.B., Goodyer, C.G., and Wang, R. (2008). Transcription factor expression in the developing human fetal endocrine pancreas. Diabetologia *51*, 1169–1180.

Maaten, L.V.D., and Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research *9*, 2579–2605.

Machado, J.D., Díaz-Vera, J., Domínguez, N., Alvarez, C.M., Pardo, M.R., and Borges, R. (2010). Chromogranins A and B as regulators of vesicle cargo and exocytosis. Cell. Mol. Neurobiol. *30*, 1181–1187.

Magenheim, J., Ilovich, O., Lazarus, A., Klochendler, A., Ziv, O., Werman, R., Hija, A., Cleaver, O., Mishani, E., Keshet, E., et al. (2011). Blood vessels restrain pancreas branching, differentiation and growth. Development *138*, 4743–4752.

Majesky, M.W., Dong, X.R., Regan, J.N., Hoglund, V.J., and Schneider, M. (2011). Vascular Smooth Muscle Progenitor Cells. Circ. Res. *108*, 365–377.

Masjkur, J., Poser, S.W., Nikolakopoulou, P., Chrousos, G., McKay, R.D., Bornstein, S.R., Jones, P.M., and Androutsellis-Theotokis, A. (2016). Endocrine Pancreas Development and Regeneration: Noncanonical Ideas From Neural Stem Cell Biology. Diabetes *65*, 314–330.

Mastracci, T.L., Anderson, K.R., Papizan, J.B., and Sussel, L. (2013). Regulation of Neurod1 contributes to the lineage potential of Neurogenin3+ endocrine precursor cells in the pancreas. PLoS Genet. *9*, e1003278.

Matsuoka, T.-A., Artner, I., Henderson, E., Means, A., Sander, M., and Stein, R. (2004). The MafA transcription factor appears to be responsible for tissue-specific expression of insulin. Pnas *101*, 2930–2933.

Maurer, P., Rorive, S., de Kerchove d'Exaerde, A., Schiffmann, S.N., Salmon, I., and de Launoit, Y. (2004). The Ets transcription factor Fev is specifically expressed in the human central serotonergic neurons. Neurosci. Lett. *357*, 215–218.

Mawe, G.M., and Hoffman, J.M. (2013). Serotonin signalling in the gut--functions, dysfunctions and therapeutic targets. Nat Rev Gastroenterol Hepatol *10*, 473–486.

McDavid, A., Finak, G., Chattopadyay, P.K., Dominguez, M., Lamoreaux, L., Ma, S.S., Roederer, M., and Gottardo, R. (2012). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. Bioinformatics *29*, 461–467.

Meivar-Levy, I., Zoabi, F., Nardini, G., Manevitz-Mendelson, E., Leichner, G.S., Zadok, O., Gurevich, M., Mor, E., Dima, S., Popescu, I., et al. (2019). The role of the vasculature niche on insulin-producing cells generated by transdifferentiation of adult human liver cells. Stem Cell Res Ther *10*, 53–10.

Mellitzer, G., Bonné, S., Luco, R.F., Van De Casteele, M., Lenne-Samuel, N., Collombat, P., Mansouri, A., Lee, J., Lan, M., Pipeleers, D., et al. (2006). IA1 is NGN3-dependent and essential for differentiation of the endocrine pancreas. Embo J. *25*, 1344–1352.

Millman, J.R., Xie, C., Van Dervort, A., Gürtler, M., Pagliuca, F.W., and Melton, D.A. (2016). Generation of stem cell-derived β-cells from patients with type 1 diabetes. Nat Commun *7*, 11463.

Miyatsuka, T., Kosaka, Y., Kim, H., and German, M.S. (2011). Neurogenin3 inhibits proliferation in endocrine progenitors by inducing Cdkn1a. Proc. Natl. Acad. Sci. U.S.a. *108*, 185–190.

Miyatsuka, T., Li, Z., and German, M.S. (2009). Chronology of islet differentiation revealed by temporal cell labeling. Diabetes *58*, 1863–1868.

Miyatsuka, T., Matsuoka, T.-A., Sasaki, S., Kubo, F., Shimomura, I., Watada, H., German, M.S., and Hara, M. (2014). Chronological analysis with fluorescent timer reveals unique features of newly generated β-cells. Diabetes *63*, 3388–3393.

Morin, X., Cremer, H., Hirsch, M.R., Kapur, R.P., Goridis, C., and Brunet, J.F. (1997). Defects in sensory and autonomic ganglia and absence of locus coeruleus in mice deficient for the homeobox gene Phox2a. Neuron *18*, 411–423.

Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J.P., et al. (2016). A Single-Cell Transcriptome Atlas of

the Human Pancreas. Cell Syst *3*, 385–394.e3.

Murtaugh, L.C. (2008). The what, where, when and how of Wnt/Œ≤-catenin signaling in pancreas development. Organogenesis *4*, 81–86.

Muzumdar, M.D., Tasic, B., Miyamichi, K., Li, L., and Luo, L. (2007). A global double-fluorescent Cre reporter mouse. Genesis *45*, 593–605.

Nair, G., and Hebrok, M. (2015). Islet formation in mice and men: lessons for the generation of functional insulin-producing β-cells from human pluripotent stem cells. Curr. Opin. Genet. Dev. *32*, 171–180.

Nair, G.G., Liu, J.S., Russ, H.A., Tran, S., Saxton, M.S., Chen, R., Juang, C., Li, M.-L., Nguyen, V.Q., Giacometti, S., et al. (2019). Recapitulating endocrine cell clustering in culture promotes maturation of human stem-cell-derived β cells. Nature Cell Biology *21*, 263–274.

Napolitano, T., Avolio, F., Courtney, M., Vieira, A., Druelle, N., Ben-Othman, N., Hadzic, B., Navarro, S., and Collombat, P. (2015). Pax4 acts as a key player in pancreas development and plasticity. Semin. Cell Dev. Biol. *44*, 107–114.

Naya, F.J., Huang, H.P., Qiu, Y., Mutoh, H., DeMayo, F.J., Leiter, A.B., and Tsai, M.J. (1997). Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. Genes Dev. *11*, 2323–2334.

Nishimura, W., Kondo, T., Salameh, T., Khattabi, El, I., Dodge, R., Bonner-Weir, S., and Sharma, A. (2006). A switch from MafB to MafA expression accompanies differentiation to pancreatic beta-cells. Dev. Biol. *293*, 526–539.

Nishimura, W., Rowan, S., Salameh, T., Maas, R.L., Bonner-Weir, S., Sell, S.M., and Sharma, A. (2008). Preferential reduction of beta cells derived from Pax6-MafB pathway in MafB deficient

mice. Dev. Biol. *314*, 443–456.

O'Rahilly, R., and Müller, F. (2010). Developmental stages in human embryos: revised and new measurements. Cells Tissues Organs (Print) *192*, 73–84.

Ohta, Y., Kosaka, Y., Kishimoto, N., Wang, J., Smith, S.B., Honig, G., Kim, H., Gasa, R.M., Neubauer, N., Liou, A., et al. (2011). Convergence of the insulin and serotonin programs in the pancreatic β-cell. Diabetes *60*, 3208–3216.

Okuno, Y., Ohtake, F., Igarashi, K., Kanno, J., Matsumoto, T., Takada, I., Kato, S., and Imai, Y. (2013). Epigenetic regulation of adipogenesis by PHF2 histone demethylase. Diabetes *62*, 1426–1434.

Oliveira, G., Ruggiero, E., Stanghellini, M.T.L., Cieri, N., D'Agostino, M., D'Agostino, M., Fronza, R., Lulay, C., Dionisio, F., Mastaglio, S., et al. (2015). Tracking genetically engineered lymphocytes long-term reveals the dynamics of T cell immunological memory. Sci Transl Med *7*, 317ra198–317ra198.

Pagliuca, F.W., Millman, J.R., Gürtler, M., Segel, M., Van Dervort, A., Ryu, J.H., Peterson, Q.P., Greiner, D., and Melton, D.A. (2014). Generation of functional human pancreatic β cells in vitro. Cell *159*, 428–439.

Pan, F.C., and Wright, C. (2011). Pancreas organogenesis: from bud to plexus to gland. Dev. Dyn. *240*, 530–565.

Paris, M., Wang, W.-H., Shin, M.-H., Franklin, D.S., and Andrisani, O.M. (2006). Homeodomain transcription factor Phox2a, via cyclic AMP-mediated activation, induces p27Kip1 transcription, coordinating neural progenitor cell cycle exit and differentiation. Mol. Cell. Biol. *26*, 8826–8839.

Pataskar, A., Jung, J., Smialowski, P., Noack, F., Calegari, F., Straub, T., and Tiwari, V.K.

(2016). NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. Embo J. *35*, 24–45.

Pattyn, A., Morin, X., Cremer, H., Goridis, C., and Brunet, J.F. (1997). Expression and interactions of the two closely related homeobox genes Phox2a and Phox2b during neurogenesis. Development *124*, 4065–4075.

Penko, D., Mohanasundaram, D., Sen, S., Drogemuller, C., Mee, C., Bonder, C.S., Coates, P.T.H., and Jessup, C.F. (2011). Incorporation of endothelial progenitor cells into mosaic pseudoislets. Islets *3*, 73–79.

Petri, A., Ahnfelt-Rønne, J., Frederiksen, K.S., Edwards, D.G., Madsen, D., Serup, P., Fleckner, J., and Heller, R.S. (2006). The effect of neurogenin3 deficiency on pancreatic gene expression in embryonic mice. Journal of Molecular Endocrinology *37*, 301–316.

Pictet, R.L., Clark, W.R., Williams, R.H., and Rutter, W.J. (1972). An ultrastructural analysis of the developing embryonic pancreas. Dev. Biol. *29*, 436–467.

Pierreux, C.E., Cordi, S., Hick, A.-C., Achouri, Y., Ruiz de Almodovar, C., Prévot, P.-P., Courtoy, P.J., Carmeliet, P., and Lemaigre, F.P. (2010). Epithelial: Endothelial cross-talk regulates exocrine differentiation in developing pancreas. Dev. Biol. *347*, 216–227.

Pinney, S.E., Oliver-Krasinski, J., Ernst, L., Hughes, N., Patel, P., Stoffers, D.A., Russo, P., and De León, D.D. (2011). Neonatal diabetes and congenital malabsorptive diarrhea attributable to a novel mutation in the human neurogenin-3 gene coding sequence. The Journal of Clinical Endocrinology & Metabolism *96*, 1960–1965.

Piper Hanley, K., Hearn, T., Berry, A., Carvell, M.J., Patch, A.-M., Williams, L.J., Sugden, S.A., Wilson, D.I., Ellard, S., and Hanley, N.A. (2010). In vitro expression of NGN3 identifies RAB3B

as the predominant Ras-associated GTP-binding protein 3 family member in human islets. J. Endocrinol. *207*, 151–161.

Piper, K., Brickwood, S., Turnpenny, L.W., Cameron, I.T., Ball, S.G., Wilson, D.I., and Hanley, N.A. (2004). Beta cell differentiation during early human pancreas development. Journal of Endocrinology *181*, 11–23.

Prado, C.L., Pugh-Bernard, A.E., Elghazi, L., Sosa-Pineda, B., and Sussel, L. (2004). Ghrelin cells replace insulin-producing beta cells in two mouse models of pancreas development. Proc. Natl. Acad. Sci. U.S.a. *101*, 2924–2929.

Prakash, N., Puelles, E., Freude, K., Trümbach, D., Omodei, D., Di Salvio, M., Sussel, L., Ericson, J., Sander, M., Simeone, A., et al. (2009). Nkx6-1 controls the identity and fate of red nucleus and oculomotor neurons in the mouse midbrain. Development *136*, 2545–2555.

Prentki, M., and Nolan, C.J. (2006). Islet beta cell failure in type 2 diabetes. J. Clin. Invest. *116*, 1802–1812.

Puri, S., and Hebrok, M. (2007). Dynamics of embryonic pancreas development using real-time imaging. Dev. Biol. *306*, 82–93.

Qi, Y., Cai, J., Wu, Y., Wu, R., LEE, J., Fu, H., Rao, M., Sussel, L., Rubenstein, J., and Qiu, M. (2001). Control of oligodendrocyte differentiation by the Nkx2.2 homeodomain transcription factor. Development *128*, 2723–2733.

Qiu, W.-L., Zhang, Y.-W., Feng, Y., Li, L.-C., Yang, L., and Xu, C.-R. (2017a). Deciphering Pancreatic Islet β Cell and α Cell Maturation Pathways and Characteristic Features at the Single-Cell Level. Cell Metab. *25*, 1194–1205.e1194.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017b).

Reversed graph embedding resolves complex single-cell trajectories. Nature Methods 2017 14:10 *14*, 979–982.

Que, J., Wilm, B., Hasegawa, H., Wang, F., Bader, D., and Hogan, B.L.M. (2008). Mesothelium contributes to vascular smooth muscle and mesenchyme during lung development. Proc. Natl. Acad. Sci. U.S.a. *105*, 16626–16630.

Reboll, M.R., Korf-Klingebiel, M., Klede, S., Polten, F., Brinkmann, E., Reimann, I., Schönfeld, H.-J., Bobadilla, M., Faix, J., Kensah, G., et al. (2017). EMC10 (Endoplasmic Reticulum Membrane Protein Complex Subunit 10) Is a Bone Marrow-Derived Angiogenic Growth Factor Promoting Tissue Repair After Myocardial Infarction. Circulation *136*, 1809–1823.

Reinert, R.B., Brissova, M., Shostak, A., Pan, F.C., Poffenberger, G., Cai, Q., Hundemer, G.L., Kantz, J., Thompson, C.S., Dai, C., et al. (2013). Vascular endothelial growth factor-a and islet vascularization are necessary in developing, but not adult, pancreatic islets. Diabetes *62*, 4154–4164.

Reinert, R.B., Cai, Q., Hong, J.-Y., Plank, J.L., Aamodt, K., Prasad, N., Aramandla, R., Dai, C., Levy, S.E., Pozzi, A., et al. (2014). Vascular endothelial growth factor coordinates islet innervation via vascular scaffolding. Development *141*, 1480–1491.

Rezania, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. Nat. Biotechnol. *32*, 1121–1133.

Rezania, A., Riedel, M.J., Wideman, R.D., Karanu, F., Ao, Z., Warnock, G.L., and Kieffer, T.J. (2011). Production of functional glucagon-secreting α-cells from human embryonic stem cells. Diabetes *60*, 239–247.

Rhim, A.D., and Stanger, B.Z. (2010). Molecular Biology of Pancreatic Ductal Adenocarcinoma Progression: Aberrant Activation of Developmental Pathways. Progress in Molecular Biology and Translational Science *97*, 41–78.

Rubio-Cabezas, O., Minton, J.A.L., Kantor, I., Williams, D., Ellard, S., and Hattersley, A.T. (2010). Homozygous mutations in NEUROD1 are responsible for a novel syndrome of permanent neonatal diabetes and neurological abnormalities. Diabetes *59*, 2326–2331.

Russ, H.A., Parent, A.V., Ringler, J.J., Hennings, T.G., Nair, G.G., Shveygert, M., Guo, T., Puri, S., Haataja, L., Cirulli, V., et al. (2015). Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. Embo J. *34*, 1759–1772.

Salisbury, R.J., Blaylock, J., Berry, A.A., Jennings, R.E., De Krijger, R., Piper Hanley, K., and Hanley, N.A. (2014). The window period of NEUROGENIN3 during human gestation. Islets *6*, e954436.

Sand, F.W., Hörnblad, A., Johansson, J.K., Lorén, C., Edsbagge, J., Ståhlberg, A., Magenheim, J., Ilovich, O., Mishani, E., Dor, Y., et al. (2011). Growth-limiting role of endothelial cells in endoderm development. Dev. Biol. *352*, 267–277.

Sander, M., Sussel, L., Conners, J., Scheel, D., Kalamaras, J., Cruz, Dela, F., Schwitzgebel, V., Hayes-Jordan, A., and German, M. (2000). Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of beta-cell formation in the pancreas. Development *127*, 5533–5540.

Sarkar, S.A., Kobberup, S., Wong, R., Lopez, A.D., Quayum, N., Still, T., Kutchma, A., Jensen, J.N., Gianani, R., Beattie, G.M., et al. (2008). Global gene expression profiling and histochemical analysis of the developing human fetal pancreas. Diabetologia *51*, 285–297.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction

of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Scavuzzo, M.A., Hill, M.C., Chmielowiec, J., Yang, D., Teaw, J., Sheng, K., Kong, Y., Bettini, M., Zong, C., Martin, J.F., et al. (2018). Endocrine lineage biases arise in temporally distinct endocrine progenitors during pancreatic morphogenesis. Nat Commun *9*, 3356.

Schaffer, A.E., Freude, K.K., Nelson, S.B., and Sander, M. (2010). Nkx6 transcription factors and Ptf1a function as antagonistic lineage determinants in multipotent pancreatic progenitors. Dev. Cell *18*, 1022–1029.

Schaffer, A.E., Taylor, B.L., Benthuysen, J.R., Liu, J., Thorel, F., Yuan, W., Jiao, Y., Kaestner, K.H., Herrera, P.L., Magnuson, M.A., et al. (2013). Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic Beta cell identity. PLoS Genet. *9*, e1003274.

Schonhoff, S.E., Giel-Moloney, M., and Leiter, A.B. (2004). Neurogenin 3-expressing progenitor cells in the gastrointestinal tract differentiate into both endocrine and non-endocrine cell types. Dev. Biol. *270*, 443–454.

Scott, M.M., Wylie, C.J., Lerch, J.K., Murphy, R., Lobur, K., Herlitze, S., Jiang, W., Conlon, R.A., Strowbridge, B.W., and Deneris, E.S. (2005). A genetic approach to access serotonin neurons for in vivo and in vitro studies. Proc. Natl. Acad. Sci. U.S.a. *102*, 16472–16477.

Segerstolpe, Å., Palasantza, A., Eliasson, P., Andersson, E.-M., Andréasson, A.-C., Sun, X., Picelli, S., Sabirsh, A., Clausen, M., Bjursell, M.K., et al. (2016). Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. Cell Metab. *24*, 593–607.

Serafimidis, I., Rodriguez-Aznar, E., Lesche, M., Yoshioka, K., Takuwa, Y., Dahl, A., Pan, D., and Gavalas, A. (2017). Pancreas lineage allocation and specification are regulated by

sphingosine-1-phosphate signalling. PLoS Biol. *15*, e2000949.

Seymour, P.A., Freude, K.K., Tran, M.N., Mayes, E.E., Jensen, J., Kist, R., Scherer, G., and Sander, M. (2007). SOX9 is required for maintenance of the pancreatic progenitor cell pool. Proc. Natl. Acad. Sci. U.S.a. *104*, 1865–1870.

Shang, Y., Yoshida, T., Amendt, B.A., Martin, J.F., and Owens, G.K. (2008). Pitx2 is functionally important in the early stages of vascular smooth muscle cell differentiation. The Journal of Cell Biology *181*, 461–473.

Shapiro, A.M.J., Pokrywczynska, M., and Ricordi, C. (2016). Clinical pancreatic islet transplantation. Nat Rev Endocrinol *13*, 268–277.

Shapiro, A.M., Lakey, J.R., Ryan, E.A., Korbutt, G.S., Toth, E., Warnock, G.L., Kneteman, N.M., and Rajotte, R.V. (2000). Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. N. Engl. J. Med. *343*, 230–238.

Sharon, N., Chawla, R., Mueller, J., Vanderhooft, J., Whitehorn, L.J., Rosenthal, B., Gürtler, M., Estanboulieh, R.R., Shvartsman, D., Gifford, D.K., et al. (2019a). A Peninsular Structure Coordinates Asynchronous Differentiation with Morphogenesis to Generate Pancreatic Islets. Cell *176*, 790–804.e13.

Sharon, N., Vanderhooft, J., Straubhaar, J., Mueller, J., Chawla, R., Zhou, Q., Engquist, E.N., Trapnell, C., Gifford, D.K., and Melton, D.A. (2019b). Wnt Signaling Separates the Progenitor and Endocrine Compartments during Pancreas Development. Cell Rep *27*, 2281–2291.e2285.

Shih, H.P., Kopp, J.L., Sandhu, M., Dubois, C.L., Seymour, P.A., Grapin-Botton, A., and Sander, M. (2012). A Notch-dependent molecular circuitry initiates pancreatic endocrine and ductal cell differentiation. Development *139*, 2488–2499.

Shih, H.P., Wang, A., and Sander, M. (2013). Pancreas Organogenesis: From Lineage Determination to Morphogenesis. Annu. Rev. Cell Dev. Biol. *29*, 81–105.

Shin, M.-H., Mavila, N., Wang, W.-H., Vega Alvarez, S., Hall, M.C., and Andrisani, O.M. (2009). Time-dependent activation of Phox2a by the cyclic AMP pathway modulates onset and duration of p27Kip1 transcription. Mol. Cell. Biol. *29*, 4878–4890.

Simon-Areces, J., Membrive, G., Garcia-Fernandez, C., Garcia-Segura, L.M., and Arevalo, M.-A. (2010). Neurogenin 3 cellular and subcellular localization in the developing and adult hippocampus. J. Comp. Neurol. *518*, 1814–1824.

Slack, J.M. (1995). Developmental biology of the pancreas. Development *121*, 1569–1580.

Smith, K.R., Kopeikina, K.J., Fawcett-Patel, J.M., Leaderbrand, K., Gao, R., Schürmann, B., Myczek, K., Radulovic, J., Swanson, G.T., and Penzes, P. (2014). Psychiatric risk factor ANK3/ankyrin-G nanodomains regulate the structure and function of glutamatergic synapses. Neuron *84*, 399–415.

Smith, S.B., Qu, H.-Q., Taleb, N., Kishimoto, N.Y., Scheel, D.W., Lu, Y., Patch, A.-M., Grabs, R., Wang, J., Lynn, F.C., et al. (2010). Rfx6 directs islet formation and insulin production in mice and humans. Nature *463*, 775–780.

Sneddon, J.B., Borowiak, M., and Melton, D.A. (2012). Self-renewal of embryonic-stem-cell-derived progenitors by organ-matched mesenchyme. Nature *491*, 765–768.

Sneddon, J.B., Tang, Q., Stock, P., Bluestone, J.A., Roy, S., Desai, T., and Hebrok, M. (2018). Stem Cell Therapies for Treating Diabetes: Progress and Remaining Challenges. Cell Stem Cell *22*, 810–823.

Sojoodi, M., Stradiot, L., Tanaka, K., Heremans, Y., Leuckx, G., Besson, V., Staels, W., Van De

Casteele, M., Marazzi, G., Sassoon, D., et al. (2016). The zinc finger transcription factor PW1/PEG3 restrains murine beta cell cycling. Diabetologia *59*, 1474–1479.

Solar, M., Cardalda, C., Houbracken, I., Mart n, M., Maestro, M.A., De Medts, N., Xu, X., Grau, V., Heimberg, H., Bouwens, L., et al. (2009). Pancreatic Exocrine Duct Cells Give Rise to Insulin-Producing ? Cells during Embryogenesis but Not after Birth. Dev. Cell *17*, 849–860.

Solomon, B.D., Pineda-Alvarez, D.E., Balog, J.Z., Hadley, D., Gropman, A.L., Nandagopal, R., Han, J.C., Hahn, J.S., Blain, D., Brooks, B., et al. (2009). Compound heterozygosity for mutations in PAX6 in a patient with complex brain anomaly, neonatal diabetes mellitus, and microophthalmia. Am. J. Med. Genet. A *149A*, 2543–2546.

Solomou, A., Meur, G., Bellomo, E., Hodson, D.J., Tomas, A., Li, S.M., Philippe, E., Herrera, P.L., Magnan, C., and Rutter, G.A. (2015). The Zinc Transporter Slc30a8/ZnT8 Is Required in a Subpopulation of Pancreatic α-Cells for Hypoglycemia-induced Glucagon Secretion. J. Biol. Chem. *290*, 21432–21442.

Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G., and Gruss, P. (1997). The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. Nature *386*, 399–402.

Soyer, J., Flasse, L., Raffelsberger, W., Beucher, A., Orvain, C., Peers, B., Ravassard, P., Vermot, J., Voz, M.L., Mellitzer, G., et al. (2010). Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development. Development *137*, 203–212.

Speer, M.Y., Yang, H.-Y., Brabb, T., Leaf, E., Look, A., Lin, W.-L., Frutkin, A., Dichek, D., and Giachelli, C.M. (2009). Smooth Muscle Cells Give Rise to Osteochondrogenic Precursors and Chondrocytes in Calcifying Arteries. Circ. Res. *104*, 733–741.

Spencer, W.C., and Deneris, E.S. (2017). Regulatory Mechanisms Controlling Maturation of Serotonin Neuron Identity and Function. Front Cell Neurosci *11*, 215.

St-Onge, L., Sosa-Pineda, B., Chowdhury, K., Mansouri, A., and Gruss, P. (1997). Pax6 is required for differentiation of glucagon-producing alpha-cells in mouse pancreas. Nature *387*, 406–409.

Stainier, D.Y.R. (2002). A glimpse into the molecular entrails of endoderm formation. Genes Dev. *16*, 893–907.

Stanescu, D.E., Yu, R., Won, K.-J., and Stoffers, D.A. (2017). Single cell transcriptomic profiling of mouse pancreatic progenitors. Physiol. Genomics *49*, 105–114.

Stoffers, D.A., Zinkin, N.T., Stanojevic, V., Clarke, W.L., and Habener, J.F. (1997). Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. Nat. Genet. *15*, 106–110.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

Su, Y., Jono, H., Misumi, Y., Senokuchi, T., Guo, J., Ueda, M., Shinriki, S., Tasaki, M., Shono, M., Obayashi, K., et al. (2012). Novel function of transthyretin in pancreatic alpha cells. FEBS Lett. *586*, 4215–4222.

Suissa, Y., Magenheim, J., Stolovich-Rain, M., Hija, A., Collombat, P., Mansouri, A., Sussel, L., Sosa-Pineda, B., McCracken, K., Wells, J.M., et al. (2013). Gastrin: a distinct fate of neurogenin3 positive progenitor cells in the embryonic pancreas. PLoS ONE *8*, e70397.

Sussel, L., Kalamaras, J., Hartigan-O'Connor, D.J., Meneses, J.J., Pedersen, R.A., Rubenstein,

J.L., and German, M.S. (1998). Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic beta cells. Development *125*, 2213–2221.

Teratani-Ota, Y., Yamamizu, K., Piao, Y., Sharova, L., Amano, M., Yu, H., Schlessinger, D., Ko, M.S.H., and Sharov, A.A. (2016). Induction of specific neuron types by overexpression of single transcription factors. In Vitro Cell. Dev. Biol. Anim. *52*, 961–973.

Terry, N.A., Walp, E.R., Lee, R.A., Kaestner, K.H., and May, C.L. (2014). Impaired enteroendocrine development in intestinal-specific Islet1 mouse mutants causes impaired glucose homeostasis. Am. J. Physiol. Gastrointest. Liver Physiol. *307*, G979–G991.

Tiveron, M.C., Hirsch, M.R., and Brunet, J.F. (1996). The expression pattern of the transcription factor Phox2 delineates synaptic pathways of the autonomic nervous system. J. Neurosci. *16*, 7649–7660.

Tokarz, V.L., MacDonald, P.E., and Klip, A. (2018). The cell biology of systemic insulin function. The Journal of Cell Biology *217*, 2273–2289.

Trikkalinou, A., Papazafiropoulou, A.K., and Melidonis, A. (2017). Type 2 diabetes and quality of life. World J Diabetes *8*, 120–129.

Tusi, B.K., Wolock, S.L., Weinreb, C., Hwang, Y., Hidalgo, D., Zilionis, R., Waisman, A., Huh, J.R., Klein, A.M., and Socolovsky, M. (2018). Population snapshots predict early haematopoietic and erythroid hierarchies. Nature Publishing Group *555*, 54–60.

Veres, A., Faust, A.L., Bushnell, H.L., Engquist, E.N., Kenty, J.H.-R., Harb, G., Poh, Y.-C., Sintov, E., Gürtler, M., Pagliuca, F.W., et al. (2019). Charting cellular identity during human in vitro β-cell differentiation. Nature *569*, 368–373.

Villasenor, A., and Cleaver, O. (2012). Crosstalk between the developing pancreas and its blood

vessels: an evolving dialog. Semin. Cell Dev. Biol. *23*, 685–692.

Villasenor, A., Chong, D.C., and Cleaver, O. (2008). Biphasic Ngn3 expression in the developing pancreas. Dev. Dyn. *237*, 3270–3279.

Villasenor, A., Chong, D.C., Henkemeyer, M., and Cleaver, O. (2010). Epithelial dynamics of pancreatic branching morphogenesis. Development *137*, 4295–4305.

Walther, D.J., and Bader, M. (2003). A unique central tryptophan hydroxylase isoform. Biochem. Pharmacol. *66*, 1673–1680.

Wang, F., Flanagan, J., Su, N., Wang, L.-C., Bui, S., Nielson, A., Wu, X., Vo, H.-T., Ma, X.-J., and Luo, Y. (2012). RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. J Mol Diagn *14*, 22–29.

Wang, H., Brun, T., Kataoka, K., Sharma, A.J., and Wollheim, C.B. (2007). MAFA controls genes implicated in insulin biosynthesis and secretion. Diabetologia *50*, 348–358.

Wang, J., Cortina, G., Wu, S.V., Tran, R., Cho, J.-H., Tsai, M.-J., Bailey, T.J., Jamrich, M., Ament, M.E., Treem, W.R., et al. (2006). Mutant neurogenin-3 in congenital malabsorptive diarrhea. N. Engl. J. Med. *355*, 270–280.

Wang, L.-L., Gu, H., Fan, Y., Zhang, Y., Wu, D., Miao, J.-N., Huang, T.-C., Li, H., and Yuan, Z.-W. Up-regulated FHL1 Expression Maybe Involved in the Prognosis of Hirschsprung's Disease. Int. J. Med. Sci. *11*, 262–267.

Wang, S., Yan, J., Anderson, D.A., Xu, Y., Kanal, M.C., Cao, Z., Wright, C.V.E., and Gu, G. (2010a). Neurog3 gene dosage regulates allocation of endocrine and exocrine cell fates in the developing mouse pancreas. Dev. Biol. *339*, 26–37.

Wang, Y.-C., Zuraek, M.B., Kosaka, Y., Ota, Y., German, M.S., Deneris, E.S., Bergsland, E.K., Donner, D.B., Warren, R.S., and Nakakura, E.K. (2010b). The ETS oncogene family transcription factor FEV identifies serotonin-producing cells in normal and neoplastic small intestine. Endocr. Relat. Cancer *17*, 283–291.

Weinreb, C., Wolock, S., and Klein, A.M. (2018). SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. Bioinformatics *34*, 1246–1248.

Weinstein, L.S., Xie, T., Qasem, A., Wang, J., and Chen, M. (2010). The role of GNAS and other imprinted genes in the development of obesity. Int J Obes (Lond) *34*, 6–17.

Wilcox, C.L., Terry, N.A., Walp, E.R., Lee, R.A., and May, C.L. (2013). Pancreatic α-cell specific deletion of mouse Arx leads to α-cell identity loss. PLoS ONE *8*, e66214.

Wilm, B., Ipenberg, A., Hastie, N.D., Burch, J.B.E., and Bader, D.M. (2005). The serosal mesothelium is a major source of smooth muscle cells of the gut vasculature. Development *132*, 5317–5328.

Winters, N., and Bader, D. (2013). Development of the Serosal Mesothelium. Jdb *1*, 64–81.

Wong, D.M., Shen, Z., Owyang, K.E., and Martinez-Agosto, J.A. (2014). Insulin- and warts-dependent regulation of tracheal plasticity modulates systemic larval growth during hypoxia in Drosophila melanogaster. PLoS ONE *9*, e115297.

Wyler, S.C., Spencer, W.C., Green, N.H., Rood, B.D., Crawford, L., Craige, C., Gresch, P., McMahon, D.G., Beck, S.G., and Deneris, E. (2016). Pet-1 Switches Transcriptional Targets Postnatally to Regulate Maturation of Serotonin Neuron Excitability. J. Neurosci. *36*, 1758–1774.

Xu, E.E., Krentz, N.A.J., Tan, S., Chow, S.Z., Tang, M., Nian, C., and Lynn, F.C. (2015). SOX4

cooperates with neurogenin 3 to regulate endocrine pancreas formation in mouse models. Diabetologia *58*, 1013–1023.

Yin, Y., Wang, F., and Ornitz, D.M. (2011). Mesothelial- and epithelial-derived FGF9 have distinct functions in the regulation of lung development. Development *138*, 3169–3177.

Yoon, J.-W., and Jun, H.-S. (2005). Autoimmune destruction of pancreatic beta cells. Am J Ther *12*, 580–591.

Yoshitomi, H., and Zaret, K.S. (2004). Endothelial cell interactions initiate dorsal pancreas development by selectively inducing the transcription factor Ptf1a. Development *131*, 807–817.

Yu, X.-X., Qiu, W.-L., Yang, L., Zhang, Y., He, M.-Y., Li, L.-C., and Xu, C.-R. (2019). Defining multistep cell fate decision pathways during pancreatic development at single-cell resolution. Embo J. *38*, e100164.

Zeng, C., Mulas, F., Sui, Y., Guan, T., Miller, N., Tan, Y., Liu, F., Jin, W., Carrano, A.C., Huising, M.O., et al. (2017). Pseudotemporal Ordering of Single Cells Reveals Metabolic Control of Postnatal β Cell Proliferation. Cell Metab. *25*, 1160–1175.e11.

Zhang, H., Ables, E.T., Pope, C.F., Washington, M.K., Hipkens, S., Means, A.L., Path, G., Seufert, J., Costa, R.H., Leiter, A.B., et al. (2009). Multiple, temporal-specific roles for HNF6 in pancreatic endocrine and ductal differentiation. Mechanisms of Development *126*, 958–973.

Zhang, Y., Deng, R., Yang, X., Xu, W., Liu, Y., Li, F., Zhang, J., Tang, H., Ji, X., Bi, Y., et al. (2017). Glucose potentiates β-cell function by inducing Tph1 expression in rat islets. Faseb J *31*, 5342–5355.

Zhao, L., Guo, M., Matsuoka, T.-A., Hagman, D.K., Parazzoli, S.D., Poitout, V., and Stein, R. (2005). The islet beta cell-enriched MafA activator is a key regulator of insulin gene

transcription. Journal of Biological Chemistry *280*, 11887–11894.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat Commun *8*, 14049.

Zheng, Y., Ley, S.H., and Hu, F.B. (2018). Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol *14*, 88–98.

Zhou, Q., Law, A.C., Rajagopal, J., Anderson, W.J., Gray, P.A., and Melton, D.A. (2007). A multipotent progenitor domain guides pancreatic organogenesis. Dev. Cell *13*, 103–114.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Daniel Wong
Author Signature

8-26-19
Date